

**DAS VIRUS TTV1
VON THERMOPROTEUS TENAX:**

**STRUKTUR, FUNKTION
UND VARIABILITÄT
EINES ARCHAEBAKTERIELLEN
GENOMS**

Horst Neumann

Disbui. München

1988



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41554823_0109

DAS VIRUS TTV1
VON *THERMOPROTEUS TENAX*:
STRUKTUR, FUNKTION UND VARIABILITÄT
EINES ARCHAEBAKTERIELLEN GENOMS

Dissertation der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

von Horst Neumann
aus Kaufbeuren

München, 1988

1. Berichterstatter: Prof. Dr. Horst Schmieger

2. Berichterstatter: Prof. Dr. August Böck

Sondergutachter: Prof. Dr. Wolfram Zillig

Tag der mündlichen Prüfung: 14. Juli 1988

Die vorliegende Arbeit wurde von Mai 1983 bis Dezember 1987 auf Anregung und unter Leitung von Herrn Professor Dr. Wolfram Zillig am Max-Planck-Institut in Martinsried durchgeführt.

Herrn Professor Zillig danke ich sehr herzlich für die großzügige Förderung der Arbeit.

Der Max-Planck-Gesellschaft danke ich für ein Promotionsstipendium.

Weiterhin danke ich Volker Schwass für technische Assistenz und für die Unterstützung bei der Auswertung von Sequenzdaten.

INHALTSÜBERSICHT

	Seite
I. ZUSAMMENFASSUNG	1
II. EINLEITUNG	4
III. MATERIAL UND METHODEN	9
Material	9
Besondere Chemikalien	9
Enzyme	10
Organismen	11
Allgemein verwendete Puffer und Lösungen	11
Methoden	12
Zucht der Bakterien	12
Zucht der Viren	13
Titerbestimmung	13
Reinigung der Viren	14
Isolierung von DNA	15
Restriktionsanalyse	15
Spaltung von DNA und Gelelektrophorese	15
Isolierung von Fragmenten zur Nachspaltung	15
Terminale Markierung von DNA	16
Southern-Transfer	17
Nick-Translation	17
Filter-Hybridisierung	18
Autoradiographie	19
Gewinnung von DNA aus Agarosegelen	19
Schmelztemperatur	20

Klonierungen	20
Klonierung in pBR 322	20
Klonierung in pIN III	20
Klonierung in M13	21
DNA-Sequenzanalyse	21
Sequenzierung mit Bal31	21
SDS-Polyacrylamid-Gel-Elektrophorese	23
Immun-Blotting	23
Gewinnung von Antikörpern	24
Andere immunologische Methoden	24
Peptidsynthese	24
Proteinsequenzanalyse	24
IV. ERGEBNISSE	26
Vermehrung und allgemeine Eigenschaften des Virus TTV1	26
Vermehrung	26
Wechselwirkung mit dem Wirt	26
Morphologie	28
Die Virusproteine	30
Die <i>Thermoproteus tenax</i> Viren TTV2, TTV3 und TTV4	30
Das Genom von TTV1	32
Elektronenmikroskopische Analyse	32
GC-Gehalt	32
Analyse des Genoms mit Restriktionsenzymen	32
Die Restriktionskarte	36

Homologie zwischen der DNA des Virus und dem Genom von <i>Thermoproteus tenax</i>	38
DNA-Sequenzanalyse	40
Offene Leserahmen auf der TTV1-DNA	47
Analyse der Strukturproteine TP1, TP2 und TP3 des Virus TTV1	50
Kartierung der Gene für TP1 und TP2	50
Identifizierung der Gene für TP1, TP2 und TP3 mit Hilfe ihrer N-terminalen Aminosäuresequenz	50
Die Organisation der Gene von TP1, TP2 und TP3	54
Kodonbenutzung der Proteine TP1, TP2 und TP3	58
Ribosomenbindungsstellen	58
Mögliche Promoterelemente	61
Eigenschaften der Strukturproteine TP1, TP2 und TP3	62
Die genomischen Varianten von TTV1	64
Struktur der Variantengenome	64
Sequenzanalyse des ClaI-Fragments Nr. 4 und Nr. 6 der genomischen Variante VT3	66
Bildung der Variante VT3	66
Das Protein TPX	73
<i>in vitro</i> Synthese eines Peptids von TPX	75
Identifizierung von TPX in der Wirtszelle	75
Das Protein TPX in der Variante VT3	77
Sekundärstruktur von TPX	79
Computervarianten von TPX	79

V. DISKUSSION	82
VI. LITERATUR	95
VII. ABKÜRZUNGEN	109

I. Zusammenfassung

Über Struktur und Funktion von Genomen im Zweig der archaebakteriellen *Thermoproteales* war zu Beginn dieser Arbeit fast nichts bekannt. Ziel der vorliegenden Untersuchung war es, das neu isolierte Virus TTV1 des Archaebakteriums *Thermoproteus tenax* zu charakterisieren und vor allem sein Genom genauer zu studieren. Dabei wurden folgende Resultate erzielt:

(1) TTV1 ähnelt morphologisch den Viren TTV2, TTV3 und TTV4 und bildet mit ihnen die Familie der stäbchenförmigen *Thermoproteus tenax* Viren.

(2) Weder auf DNA-, noch auf Proteinebene konnte für diese Viren Homologie nachgewiesen werden.

(3) TTV1 kommt in den Zellen in einem nicht näher definierten Zustand vor, der sowohl zur Virusproduktion und Lyse, aber auch mit relativ großer Frequenz zur Segregation kurierter Zellklone führen kann. Das Virus-Genom ist zu keinem Zeitpunkt in das Genom des Wirts integriert.

(4) Die Vermehrung des Virus wird vom Energiezustand des Wirts beeinflusst.

(5) Das Genom des Virus ist eine 15,9 kb große, doppelsträngige, lineare DNA mit einem GC-Gehalt von ca. 37%.

(6) Die DNA des Virus ist in unbekannter Weise terminal



maskiert. Dies zeigt sich durch substöchiometrische terminale Markierung und den hydrophoben Charakter der Endfragmente der TTV1-DNA.

(7) Die DNA des Virus wurde zu 83% sequenziert. Es fanden sich dabei auf beiden Strängen insgesamt 38 offene Leserahmen von mehr als 150 bp Länge.

(8) Die Gene für die Strukturproteine TP1, TP2 und TP3 wurden kartiert, ihre genaue Lage auf dem Genom wurde identifiziert und die Eigenschaften der Proteine wurden charakterisiert.

(9) Auf der DNA wurden mögliche Promotorstrukturen, aber keine Ribosomenbindungsstellen entdeckt.

(10) Bis jetzt konnten fünf genomische Varianten von TTV1 entdeckt werden, ein Hinweis auf genetische Variabilität in *Thermoproteus tenax*.

(11) Die Genome der Varianten unterscheiden sich durch Deletionen, Insertionen und verschiedene Anordnung von kurzen DNA-Sequenzen.

(12) Zwei Restriktionsfragmente, die in die Variantenbildung involviert sind, wurden sequenziert und charakterisiert. Ein Modell für die Variantenbildung wurde entwickelt.

(13) Das Genprodukt eines offenen Leserahmens von 462 bp, das Protein TPX, wurde mit Hilfe von Antikörpern gegen ein *in vitro* synthetisiertes Peptid dieses Proteins in virustragenden *Thermoproteus tenax* Zellen nachgewiesen.

(14) Das Protein TPX wurde im Wildtyp-Virus, in der Variante VT3 und in weiteren, mit Hilfe des Computers gebildeten Varianten, charakterisiert.

II. Einleitung

Bei der Aufstellung eines phylogenetischen Systems der Organismen durch vergleichende Sequenzanalyse von Spaltstücken der 16S rRNA bzw. der 18S rRNA entdeckten vor etwa einem Jahrzehnt C. Woese und Mitarbeiter, daß sich einige Mikroorganismen sowohl von den übrigen Prokaryonten als auch von den Eukaryonten deutlich unterschieden. Sie nannten diese Gruppe von Lebewesen "Archaeobakterien" und stellten sie als drittes Urreich den beiden anderen Urreichen der Eubakterien und der Eukaryonten gegenüber (Woese & Fox, 1977; Woese *et al.*, 1978; Fox *et al.*, 1980). Phylogenetisch sind die Archaeobakterien in drei Hauptgruppen einzuordnen (Woese & Olsen, 1986; Klenk *et al.*, 1987): (1) in die Methanogenen (*Methanococcales*, *Methanobacteriales* und *Methanomicrobiales*) (Whitman, 1985) und die extrem Halophilen (*Halobacteriales* und *Thermoplasmatales*) (Kushner, 1985), (2) in die schwefelabhängigen, extrem thermophilen *Thermococcales* (Woese & Olsen, 1986; Zillig *et al.*, 1987) und (3) in die schwefelabhängigen, extrem thermophilen *Thermoproteales* und *Sulfolobales* (Stetter & Zillig, 1985).

Für die Archaeobakterien typische Merkmale sind z.B. die Natur ihrer Lipide (Langworthy *et al.*, 1972, 1974, 1977; Tornabene *et al.*, 1979a, 1979b), die Abwesenheit eines

Mureinsacculus (Kandler, 1979), die Besonderheit ihrer ribosomalen Proteine (Matheson, 1982), ihre Resistenz gegen die meisten Antibiotika, die das Wachstum der Eubakterien hemmen (Hilpert *et al.*, 1981; Pecher & Böck, 1981) und ihre extreme Lebensweise bzw. spezielle Physiologie.

Die genauere molekularbiologische Charakterisierung der Archaeobakterien hat aber auch ergeben, daß sie gewisse Merkmale mit den Eubakterien, andere aber mit den Eukaryonten gemeinsam haben.

In der Größe der Ribosomen, der Organisation von Genen in größeren Transkriptionseinheiten, der Anwesenheit von Shine-Dalgarno Sequenzen auf mRNAs (Review in Dennis, 1986) und der prokaryotischen Organisation der Zellen, ähneln sie den Eubakterien. Auf der anderen Seite sind Gemeinsamkeiten zwischen Archaeobakterien und Eukaryonten z.B. das Vorkommen von Introns in Genen für stabile RNAs (Kaine *et al.*, 1983; Daniels *et al.*, 1985; Kjems & Garrett, 1985; Wich *et al.*, 1987; Kaine, 1987), das Muster der Untereinheiten und die Aminosäuresequenz der DNA-abhängigen RNA-Polymerasen (Huet *et al.*, 1983; Zillig *et al.*, 1985a, 1985b; Pühler *et al.*, 1988), die Eigenschaften und Aminosäuresequenzen von Elongationsfaktoren des Proteinbiosyntheseapparates (Kessel & Klink, 1980; Lechner & Böck, 1987) und das Vorkommen von durch Aphidicolin hemmbaren DNA-Polymerasen (Zabel *et al.*, 1985; Prangishvili, 1986).

Die archaeobakteriellen Viren

Untersuchungen von Virus-Wirtssystemen in den anderen Urreichen waren von außerordentlicher Bedeutung für das Verständnis der Organisation und Funktion ihrer Genome, insbesondere von Signalstrukturen und Kontrollmechanismen in der Genexpression. Aus diesem Grund ist auch die Verfügbarkeit von archaeobakteriellen Viren für die vergleichende molekularbiologische Analyse des genetischen Apparates der Archaeobakterien von großem Nutzen.

Schon bevor ihre Wirte als Archaeobakterien erkannt wurden, hatte man zwei Viren von extrem halophilen Archaeobakterien isoliert, den Phagen Hs1 von *Halobacterium salinarium* (Torsvik, 1982; Torsvik & Dundas, 1974, 1978, 1980) und den *Halobacterium* Phagen Jai (Wais et al., 1975). Andere Viren dieses Zweiges der Archaeobakterien sind die Halobakterienphagen ϕ H (Schnabel et al., 1982a), Hh1 und Hh3 (Pauling, 1982), S45 (Daniels & Wais, 1984) und ϕ N (Vogelsang-Wenke, 1984) und der Phage PG von *Methanobrevibacter smithii* (Bertani & Baresi, 1986).

Bei den *Thermococcales* gibt es nur ein durch seine ikosaedrische Struktur virusähnliches Partikel in *Pyrococcus woesei* (Zillig et al., 1987), das noch nicht näher untersucht wurde.

In den *Thermoproteales* und *Sulfolobales* entdeckte man die Viren TTV1, TTV2, TTV3 und TTV4 in *Thermoproteus tenax*

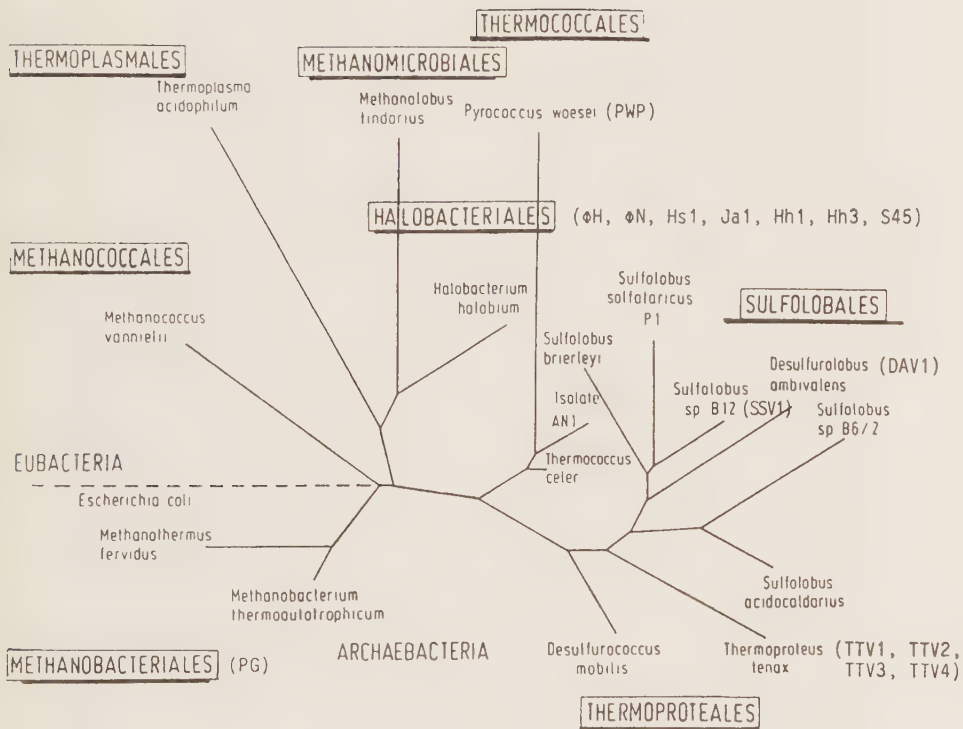


Abb. 1 : Die Viren der Archaeobakterien

(Janekovic *et al.*, 1983, Zillig *et al.*, 1986b), das virusähnliche Partikel SSV1 in *Sulfolobus sp.* B12 (Martin *et al.*, 1984) und das nicht näher untersuchte Virus DAV1 in *Desulfurolobus ambivalens* (Zillig *et al.*, 1986a).

Die vorliegende Arbeit befaßt sich mit dem Virus TTV1 von *Thermoproteus tenax*, das von D. Janekovic elektronenmikroskopisch nachgewiesen und von W. Zillig erstmals gereinigt wurde. Das Ziel der Arbeit war, anhand dieses Virus ein archaebakterielles Genom und Virus-Wirtssystem zu untersuchen. Dies geschah neben physiologischen Untersuchungen zum überwiegenden Teil durch Sequenzierung der TTV1-DNA, der Charakterisierung der offenen Leserahmen und der Zuordnung der Gene zu den Hauptproteinen des Virus. Durch die Beobachtung von Veränderungen in der Virus-DNA ergab sich als weiteres Ziel, in TTV1 möglicherweise ein ähnliches System genetischer Variabilität wie in dem *Halobacterium* Phagen ϕ H (Schnabel *et al.*, 1982a, 1982b) nachzuweisen.

Ein Teil der hier zu berichtenden Ergebnisse wurde bereits veröffentlicht (Janekovic *et al.* 1983, Zillig *et al.*, 1986b).

III. MATERIAL UND METHODEN

MATERIAL

Besondere Chemikalien

Anorganische Salze	Merck, Darmstadt
Cäsiumchlorid	Biomol, Ilvesheim
Phenol p.a.	Merck, Darmstadt
Natriumdodecylsulfat	Serva, Heidelberg
Triton-X-100	Serva, Heidelberg
Acrylamid	Serva, Heidelberg
N,N'-Methylenbisacrylamid	Serva, Heidelberg
TEMED	Serva, Heidelberg
Ammoniumperoxodisulfat	Merck, Darmstadt
2-Mercaptoethanol	Roth, Karlsruhe
Formamid p.a.	Merck, Darmstadt
Ethidiumbromid	Sigma, St.Louis
Coomassie Brilliant Blue R 250	Serva, Heidelberg
Bromphenolblau	Merck, Darmstadt
Xylencyanol	Serva, Heidelberg
Agarose	BRL, Gaithersburg
Sea Plaque Agarose	FMC, Rockland
Gelatine	Merck, Darmstadt
Sephadex G-50	Pharmacia, Uppsala
DNA, aus Lachssperma	Sigma, St.Louis
BSA	Calbiochem, La Jolla

Nitrocellulose (BA 85, 0,45 µm)	Schleicher & Schüll, Dassel
IPTG	Biomol, Ilvesheim
X-Gal	Biomol, Ilvesheim
³² P-Radionukleotide	Amersham, Braunschweig
³⁵ S-Radionukleotide	Amersham, Braunschweig
Desoxy-nukleotidtriphosphate	Sigma, St.Louis
Didesoxy-nukleotidtriphosphate	Sigma, St.Louis
Harnstoff (ultra pure)	BRL, Gaithersburg
Whatman 3 MM	Whatman, New Jersey
4-Chloro-1-naphtol	Merck, Darmstadt
Geneclean	Bio 101, La Jolla
ortho-Dianisidin	Fluka, Buchs
Nonidet	Sigma, St.Louis
Peroxidase konjugierter	
Kaninchenantikörper	Sigma, St.Louis
Freund's Adjuvant (complete)	Sigma, St.Louis
Röntgenfilm X-Ray 90	Linhard, München
Trimax Röntgenfilm	3M, Italien

Enzyme

Alkalische Phosphatase	Boehringer, Mannheim
Desoxyribonuklease I	Boehringer, Mannheim
Proteinase K	Boehringer, Mannheim
Pronase	Boehringer, Mannheim
Restriktionsendonukleasen	Boehringer, Mannheim
Bal31 Nuklease	Boehringer, Mannheim
DNA-Polymerase I	Boehringer, Mannheim

DNA-Polymerase I (Klenow Fragment)	Pharmacia, Uppsala
Polynukleotidkinase	NEN, Dreieich
T ₄ -DNA-Ligase	NEN, Dreieich

Organismen

E. coli HB101: F⁻ hsd S20 ($x^-_B m^-_B$) recA13 (Boyer und Rolland-Dussoix, 1969)

E. coli JM101: Δ (lacpro), thi, supE, F'traD36, proAB, lacI^qZAM15 (Messing, 1983)

Thermoproteus tenax Kral, DSM 2078 wurde isoliert aus einem Schlammloch (93°C, pH 6) im Solfatarenfeld der Krafla in Island (Zillig et al., 1981). Die Viren TTV1, TTV2, TTV3 und TTV4 wurden in chemolithoautotrophen Kulturen dieses Organismus gefunden (Janekovic et al., 1983).

Allgemein verwendete Puffer und Lösungen

TE-Puffer:

20 mM Tris/HCl pH 8,0

1 mM EDTA

Formamid-Probenpuffer:

Je 0,3 mg/ml Bromphenolblau und Xylencyanol sowie 20 mM Na-EDTA in entionisiertem Formamid. Die Farbstoffe wurden in fester Form, das EDTA als 0,1 M Lösung (pH 8,3) zugegeben.

Phenol zur Nukleinsäureextraktion:

Phenol mit 0,02% 8-Hydroxychinolin als Stabilisator, äquilibriert mit TE-Puffer.

20x SSC:

3 M NaCl

0,3 M Trinatriumcitrat pH 7,0

METHODEN

Zucht der Bakterien

Die Organismen wurden anaerob bei 88°C und pH 5,2 im Schüttelbad in Brock's Basalsalzmedium (Brock, 1972) gezogen. Heterotrophe Zucht wurde durchgeführt mit 0,02% Hefeextrakt, 1% Saccharose und 0,5% (w/v) elementarem Schwefel als Elektronenakzeptor. Medien für autotrophe Zucht enthielten ebenfalls 0,5% (w/v) elementaren Schwefel. In diesem Fall wurde mit H₂ (10% der Gasphase) und CO₂ (90% der Gasphase) als Kohlenstoffquelle begast.

Zucht der Viren

Große Mengen von TTV1 wurden auf zwei Arten produziert. Entweder wurde der sensitive Stamm *T. tenax* Kral/2 mit einer m.o.i. von 0,1-0,5 mit TTV1 in Anwesenheit eines Überschusses von Schwefel während der exponentiellen Wachstumsphase bei $1-2 \times 10^8$ Zellen/ml infiziert und die Viren nach der Lyse geerntet, oder der lysogene Stamm *T. tenax* Kral/5 wurde solange gezogen bis der Schwefel verbraucht war, worauf Zellyse mit anschließender Freisetzung von TTV1 eintrat.

Titerbestimmung

Da bis jetzt wegen der Schwierigkeiten, den Wirt als Zellrasen zu plattieren, noch kein Plaquetest möglich war, wurde die Vermehrung durch elektronenmikroskopische Zählung festgestellt. Proben der Lysate wurden durch Zentrifugation von den Zellen gereinigt. Das Grid wurde 5 min auf einen Tropfen des Überstandes gelegt, dann mit H_2O gewaschen, 3 min mit 2% (w/v) Glutaraldehyd fixiert, wieder mit H_2O gewaschen und in einem Balzer BAE 3701 Rotationsbedampfer mit Platin bedampft. Die Viruspartikel in 10 sichtbaren Feldern wurden bei einer Vergrößerung von 6.000 in einem ELM11 Elektronenmikroskop ausgezählt. Die Methode wurde kalibriert durch Vergleich dieser Zahl mit der absoluten Anzahl an Partikeln. Die absolute Anzahl wurde festgestellt,

indem die Viren 2 h bei 10.000 rpm in einem SW41 Rotor in einer Beckman Ultrazentrifuge auf ein Grid am Boden des Röhrchens zentrifugiert wurden. Die Probe wurde sofort fixiert durch Zugabe eines Volumens 4% (w/v) Glutaraldehyd zum Überstand. Die Abwesenheit von Viruspartikeln im Überstand wurde elektronenmikroskopisch festgestellt. Das Grid wurde danach ohne es zu waschen mit Platin bedampft und die Partikel in einem bestimmten Areal ausgezählt. Von der gemessenen Höhe der Probe über dem Grid aus wurde die absolute Konzentration an Viruspartikeln pro Milliliter der Probe kalkuliert. Ein Partikel in der Standardmethode (s.o.) entsprach demnach 7×10^6 Partikeln pro Milliliter.

Reinigung der Viren

Lysate wurden durch Hyflo Super-cel (Johns-Manville) filtriert. Die Viren wurden durch Einrühren von 40% Ammoniumsulfat gefällt. Nach Zentrifugation durch eine Padberg ZKW 320 Durchflußzentrifuge wurde das Präzipitat in einer CsCl-Lösung (2,3 M in Brock's Medium) resuspendiert und im VTi 50 Rotor bei 50.000 rpm und 15°C 24 h zentrifugiert. Die Phagenbande wurde mit einer Spritze gezapft und im VTi 65 Rotor nach Auffüllen mit der CsCl-Lösung rebandierte (10 h, 65.000 rpm, 15°C).

Isolierung von DNA

CsCl-bandierte Viren wurden gegen Brock's Salze dialysiert und nach Zugabe von SDS (1% w/v) anschließend je dreimal mit Phenol, Chloroform-Oktanol (9:1) und Chloroform ausgeschüttelt. Wenn nötig, wurde die DNA mit zwei Volumenteilen Ethanol nach Zugabe von Natriumacetat (Endkonzentration 0,3 M) gefällt.

Restriktionsanalyse

Spaltung von DNA und Elektrophorese

Die Restriktionsspaltungen wurden nach Angaben der Hersteller durchgeführt. Für die gelelektrophoretische Auftrennung von DNA-Fragmenten wurden 0,7-2%ige horizontale Agarosegele verwendet. Die Elektrophorese erfolgte in Tris-Acetat-Puffer (Maniatis *et al.*, 1982) bei 40V/80mA über Nacht. Die Gele wurden 15 min in 0,5 µg/ml Ethidiumbromid gefärbt, 30 min in Wasser entfärbt und unter UV-Licht ($\lambda=254$ nm) photographiert.

Isolierung von Fragmenten zur Nachspaltung

(nach Parker & Seed, 1980)

5 µg DNA wurden nach der Restriktionsspaltung in einem 0,7%igen Gel aus niedrigschmelzender Agarose (Sea Plaque) in

Tris-Acetat-Puffer aufgetrennt. Nach Anfärben der DNA wurde das Gel in dem Puffer äquilibriert, der für die Nachspaltung benötigt wurde (10 mM Tris-HCl pH 7,8, 6 mM MgCl_2 und NaCl nach Bedarf). Die Gelabschnitte mit den DNA-Fragmenten wurden unter langwelligem UV-Licht ($\lambda=330$ nm) ausgeschnitten, die Agarose bei 70°C geschmolzen und auf 37°C abgekühlt, wobei sie flüssig blieb. Nach Zugabe von 0,1% BSA und 3 U des zweiten Restriktionsenzymes folgte eine dreistündige Inkubationszeit. Die zweite Gelelektrophorese wurde mit auf 37°C vorgewärmtem Gel und Puffer durchgeführt, um das Erstarren der aufgetragenen Probe vor dem Eindringen der DNA in das Gel zu verhindern.

Terminale Markierung von DNA

10 μg DNA wurden mit 0,1 U alkalischer Phosphatase in 100 μl 0,1 M NaCl, 0,01 M Tris-HCl, pH 7,6 dephosphoryliert, dreimal phenolisiert und mit Ethanol gefällt. Die Markierung erfolgte in 25 μl 50 mM Tris-HCl, 10 mM MgCl_2 , 5% Glycerin, 5 mM 2-Mercaptoethanol mit 250 μCi ($\gamma\text{-}^{32}\text{P}$)-ATP und 10 U Polynucleotidkinase bei 37°C für 30 min. Nach Ausschütteln mit Phenol und Chloroform wurde unverbrauchtes ATP über eine Sephadex-G 50 Säule in TE-Puffer abgetrennt.

Southern-Transfer

Der Transfer von DNA-Fragmenten von Agarosegelen auf Nitrocellulosefilter erfolgte im Wesentlichen wie von Southern (1975) und Maniatis *et al.* (1982) beschrieben.

Nick-Translation

Zur radioaktiven Markierung von DNA durch Nick-Translation (Rigby *et al.*, 1977) wurde folgender 10 µl-Ansatz verwendet:

50 mM Tris-HCl pH 7,5

10 mM MgCl₂

10 mM 2-Mercaptoethanol

50 µg/ml Gelatine

10 µM dCTP, dGTP, dTTP

50 µCi (α -³²P)-dATP (3000 Ci/mMol)

10 pg DNAase I

3 U E. coli DNA-Polymerase I

0,2 µg zu markierende DNA

Nach 3 h Inkubation bei 14-16°C wurden 40 µl H₂O und 15 µl Stop-Lösung (50 mg/ml SDS, 2,5 mg/ml Bromphenolblau,

150 mM EDTA) hinzugefügt und die nicht eingebauten Triphosphate über eine Sephadex-G 50 Säule abgetrennt.

Filter-Hybridisierung

Gebackene Nitrocellulosefilter mit gebundener DNA wurden in 5x SSC angefeuchtet, in verschweißbaren Tüten 4 h bei 68°C vorhybridisiert und anschließend 16 h bei 42°C mit radioaktiv markierter DNA hybridisiert (ca. 5.000 cpm pro cm² Filterfläche). Die Filter wurden anschließend 15 min bei Raumtemperatur in Waschlösung geschüttelt und das Waschen dann für 2-4 h bei höherer Temperatur (60-65°C) fortgesetzt. Für die Autoradiographie wurden die noch feuchten Filter auf mit Waschlösung durchtränktes Whatman 3MM-Papier gelegt und mit Haushaltsfolie abgedeckt. Als Unterlage für das Whatman 3MM-Papier dienten nicht mehr benötigte entwickelte Röntgenfilme.

Lösungen

Prähybridisierungslösung

5x SSC

0,5% SDS

200 µg/ml Lachssperma-DNA (hitzedenaturiert)

Hybridisierungslösung

50% Formamid p.A.

5x SSC

0,2% SDS

1/20 V.T. radioaktiv markierte Probe

Waschlösung

2x SSC

0,1% SDS

Autoradiographie

Die Autoradiographie von Nitrocellulosefiltern oder von Sequenzgelen erfolgte bei Raumtemperatur unter Verwendung von X-ray 90 bzw. Trimax Röntgenfilm oder aber bei -80°C mit Hilfe einer Verstärkerfolie (Trimax).

Gewinnung von DNA aus Agarosegelen

DNA-Fragmente wurden durch Auflösen der Agarosestücke in Natriumjodidlösung und Adsorption der DNA an Glaspulver gewonnen (Vogelstein & Gillespie, 1979). Die Durchführung dieser Methode erfolgte mit dem GENE CLEAN-Kit der Firma BIO 101 entsprechend den Instruktionen des Herstellers.

Schmelztemperatur

Die Schmelztemperatur von TTV1-DNA wurde nach Marmur & Doty (1962) bestimmt.

Klonierungen

Klonierungen in pBR 322

Die Ligation von gereinigten DNA-Fragmenten in pBR 322 (Bolivar *et al.*, 1979) die Transformation von *E. coli* HB101 mit dem Ligationsansatz und die Identifizierung rekombinanter Klone erfolgte nach Standardverfahren (Maniatis *et al.*, 1982).

Klonierungen in pIN III

Die Klonierungen in pIN III (Inouye, 1983) erfolgten wie in pBR 322. Die Identifizierung rekombinanter Klone mit Hilfe von Antiseren wurde nach Konheiser *et al.* (1984) durchgeführt.

Klonierungen in M13

Die Klonierungen von definierten Fragmenten der TTV1-DNA in die M13-Vektoren M13tg130 und M13tg131 (Kieny *et al.*, 1983) erfolgten entsprechend einem Protokoll von Amersham International. Als Wirt diente *E. coli* JM101.

DNA-Sequenzanalyse

DNA-Sequenzen wurden nach der enzymatischen Methode (Sanger *et al.*, 1977) unter Verwendung des M13-Klonierungssystems (Sanger *et al.*, 1980; Messing & Viera, 1982) bestimmt. Die Sequenzierung erfolgte nach der Bal31-Methode von Poncz *et al.* (1982). Die Sequenzreaktionen und die Elektrophorese wurden entsprechend einem Protokoll von Amersham International durchgeführt. Zur radioaktiven Markierung wurde (^{35}S)-dATP α S (800 Ci/mol) eingesetzt. Zur Auswertung der DNA-Sequenzdaten wurden die Programme von Staden (1980), Devereux *et al.* (1984) und der Protein Identification Resource (PIR) verwendet.

Sequenzierung mit Bal31

Der Methode liegt zugrunde, daß die Nuklease Bal31 DNA von beiden Enden her abbaut.

Restriktionsfragmente der TTV1-DNA wurden im M13-System

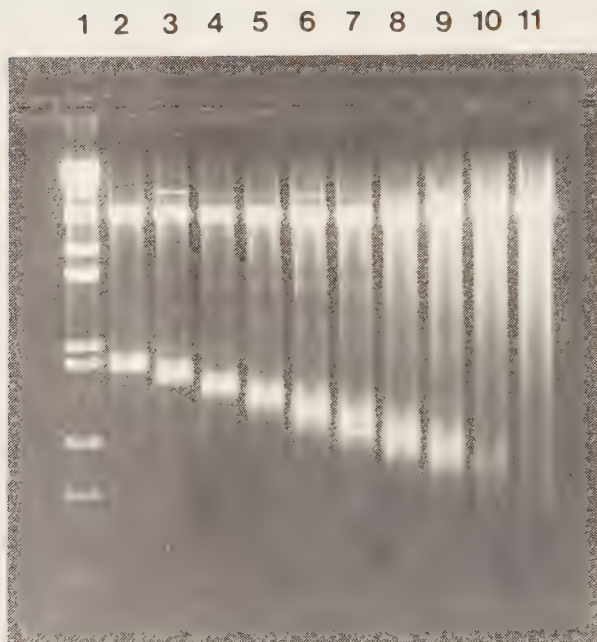


Abb. 2 : Abbau eines Restriktionsfragmentes
mit Bal31 und Auftrennung im Agarosegel.
(1) Hind III gespaltene T₅-DNA
(2)-(11) Bal31-Verdau: zu ⁵verschiedenen
Zeitpunkten entnommene DNA-Proben.

kloniert. 50 µg dieser DNA wurden mit einem Restriktionsenzym, das nahe dem Insert im Vektor spaltet, linearisiert und anschließend 10 min mit 15 U Bal31 inkubiert. Jede Minute wurde eine Probe genommen. Anschließend wurde die Nuklease durch Phenolisieren entfernt, die abgedaute Insert-DNA mit einem zweiten Restriktionsenzym vom Vektor abgespalten, über ein Agarosegel abgetrennt und mit Hilfe des GENE CLEAN-Kits aus der Agarose eluiert. Die abgedauten DNA-Stücke wurden erneut in M13 kloniert und anschließend sequenziert.

SDS-Polyacrylamid-Gel-Elektrophorese

Diese wurde wie von Mirault & Scherrer (1971) beschrieben durchgeführt. Die Virussuspensionen oder -fraktionen wurden entweder direkt in Laemmli's Probenpuffer (Laemmli, 1970) aufgenommen oder zuerst mit Phenol extrahiert. Im letzteren Fall wurden die Proteine aus der Phenolphase durch Präzipitation mit vier Volumenteilen Aceton gewonnen und anschließend in Laemmli's Probenpuffer aufgenommen.

Immun-Blotting

Das Western-Blot Verfahren wurde nach Towbin *et al.* (1979) und Burnette (1980) durchgeführt.

Gewinnung von Antikörpern

Die Elution der aufgetrennten Proteine aus dem Gel, die Immunisierung der Kaninchen und die Gewinnung von Antikörpern erfolgte nach Schnabel *et al.* (1983).

Andere immunologische Methoden

Die Gewinnung von Zellysaten für den Western-Blot erfolgte nach Huet *et al.* (1982). ELISA-Tests wurden wie bei Grant & Williams (1983) nach Voller *et al.* (1979) durchgeführt.

Peptidsynthese

Das iso-1-Cytochrom C-Konjugat mit dem Peptid β Ala-(Thr-Pro)₃-Thr-Tyr-Asp-Ile-Thr-Tyr-Val-Val-Phe-Asp-Val-Thr-Pro-Ser-Pro-Thr-ProOH wurde freundlicherweise von E. Wünsch, Martinsried, zur Verfügung gestellt. Das Peptid wurde nach Standardverfahren der Peptidsynthese hergestellt.

Proteinsequenzanalyse

Die Bestimmung der Aminosäurezusammensetzung und der Aminosäuresequenzen wurde von F. Lottspeich und R. Hensel, Martinsried, durchgeführt. Die Proteine wurden auf einem 15%igen Polyacrylamidgel aufgetrennt und elektrophoretisch

auf ein derivatisiertes Glasfiberfilter transferiert. Die Banden wurden nach Anfärben mit Coomassie Brilliant Blue R250 ausgeschnitten und direkt in einem Gasphasensequenator (Applied Biosystems 470A) sequenziert. Die Phenyl-Thiohydantoin-Aminosäuren Derivate wurden durch HPLC Analyse nach Lottspeich (1985) identifiziert.

IV. ERGEBNISSE

WACHSTUM UND ALLGEMEINE EIGENSCHAFTEN DES VIRUS TTV1

Vermehrung

In flüssiger Kultur vermehrt sich das Virus TTV1 auf *Thermoproteus tenax* trotz jeweils gleichen Bedingungen bis zu sehr unterschiedlichen Titern. Der bisher maximal erreichte Titer war 3×10^{10} Viren/ml, das entspricht bei vollständiger Lyse einer Zellpopulation von 1×10^8 Zellen/ml einer Produktion von bis zu 300 Viren pro *Thermoproteus*-zelle.

Da es bisher noch nicht möglich war, *Thermoproteus tenax* als Rasen zu plattieren (nur Einzelkolonien konnten beobachtet werden), war auch kein Plaquetest verfügbar, ein Umstand, der die genetische Arbeit stark behinderte.

Wechselwirkung mit dem Wirt

TTV1 tragende, nicht infizierte Zellen produzieren Virus in autotrophen Kulturen. Die Lyse beginnt, wenn der Schwefel durch die Bakterien verbraucht ist. Das scheint darauf hinzudeuten, daß die Lyse in einer bis jetzt unbekannten Weise durch Energiemangel ausgelöst wird.

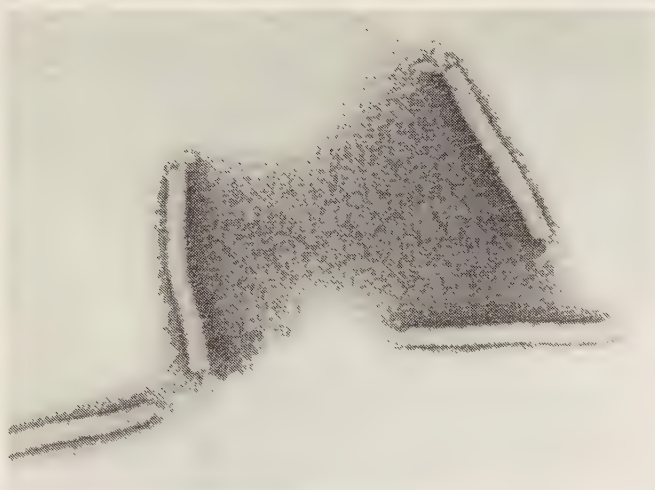


Abb. 3 : Das *Thermoproteus tenax* Virus TTV1
(Negativ-Kontrastierung).

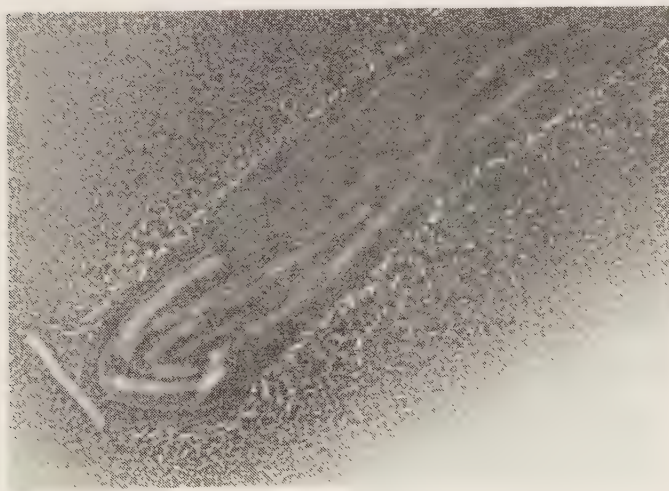


Abb. 4 : *Thermoproteus tenax* Zelle mit intrazellulären
TTV1-Viren (Negativ-Kontrastierung).

Diesen undefinierten Zusammenhang zwischen Wirt und Virus, der von unvollständiger Lyse und kurierten bzw. resistenten Zellen bis zur vollständigen Lyse reicht, könnte man im Gegensatz zur klassischen Lysogenie als eine Art "carrier state" bezeichnen. Virus-tragende Zellen enthalten das Virus-Genom in linearer, nicht in zirkulärer oder ins Wirts-Genom integrierter Form, wie durch radioaktiv markierte TTV1-DNA nachgewiesen wurde.

Durch Plattieren oder mit Hilfe von Verdünnungsreihen erhält man kurierte, virus-sensitive Klone, auf denen das Virus sowohl in autotrophen als auch in heterotrophen Kulturen auch dann produziert wird, wenn noch Schwefel vorhanden ist. Die Lyse tritt dabei zwischen 24 und 48 Stunden nach der Infektion ein, und man erhält Titer zwischen 1×10^{10} und 2×10^{10} Viruspartikeln pro Milliliter Lysat.

Morphologie

Eine elektronenmikroskopische Aufnahme des Virus TTV1 (Abb. 3) zeigt ein kurzes, stäbchenförmiges Partikel von 0,4 μm Länge und 0,04 μm Durchmesser. Es ähnelt damit eher den helicalen Pflanzenviren (wie z.B. dem Tabakmosaikvirus) als den fadenförmigen Bakteriophagen wie fd oder gar den klassischen Bakteriophagen wie λ . Auch unter den Phagen der Archaeobakterien ist die Morphologie der *Thermoproteus tenax* Viren bisher einmalig.

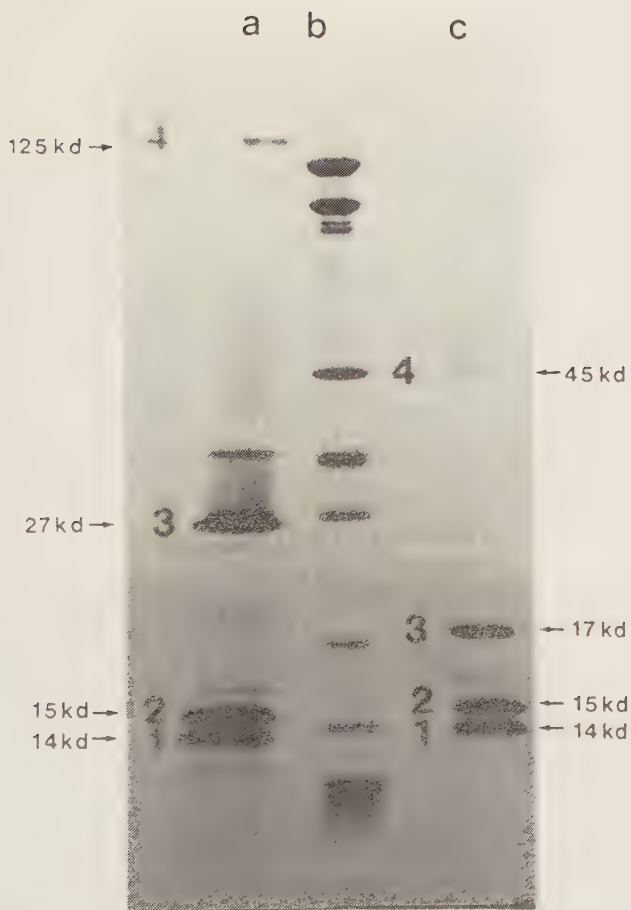


Abb. 5 : SDS-Polyacrylamid-Gel-Elektrophorese der Proteinkomponenten von TTV1.

(a) TTV1 Proteine unbehandelt

(b) *Sulfolobus* RNA-Polymerase

(c) TTV1 Proteine nach Phenolbehandlung

Die Virusproteine

Nach Auftrennung der Proteine des Virus TTV1 durch SDS-Polyacrylamid-Gel-Elektrophorese erkennt man vier Hauptproteine von 14 kd, 15 kd, 27 kd und 125 kd Größe. Die Proteine TP3 und TP4 ändern nach Phenolbehandlung ihr Laufverhalten im SDS-Polyacrylamid-Gel und bandieren nun bei Größen von 17 kd bzw. 45 kd (Abb. 5). Das läßt darauf schließen, daß die Proteine entweder jeweils mit einer bisher unbekannten Nicht-Protein-Komponente verbunden sind, die durch Phenolbehandlung entfernt wird, oder aber daß die Proteine eine Konformationsänderung erfahren und deshalb ihr Laufverhalten im SDS-Gel ändern.

Die *Thermoproteus tenax* Viren TTV2, TTV3 und TTV4

Alle diese Viren ähneln einander in ihrer Stäbchenform (Abb. 6) und in der Form ihres Genoms, einer linearen, doppelsträngigen DNA (Tab. 1). Obwohl sie alle den gleichen Wirt haben und man sie daher einer Familie, den stäbchenförmigen *Thermoproteus tenax* Viren zuordnen kann, zeigte sich durch Kreuzhybridisierungsexperimente, daß sowohl auf DNA- als auch auf Proteinebene zumindest keine enge Verwandtschaft zwischen den Viren besteht.

Virus	Morphologie	Größe des Virions (nm)	Größe der DNA (kb)
TTV1	Stäbchen	40 x 400	15,9
TTV2	flexibles Stäbchen	20 x 1250	16
TTV3	flexibles Stäbchen	30 x 2500	27
TTV4	Stäbchen	30 x 500	17

Tab. 1 : Die Viren von *Thermoproteus tenax*

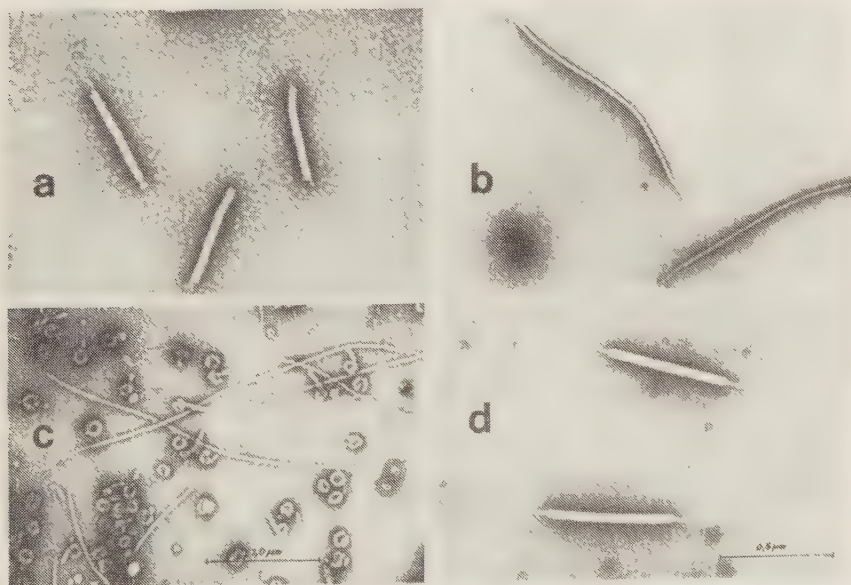


Abb. 6 : Die Viren von *Thermoproteus tenax*
 (a) TTV1, (b) TTV2, (3) TTV3 mit Überresten
 von TTV1, (d) TTV4.

DAS GENOM VON TTV1

Elektronenmikroskopische Analyse

Das Genom des Virus ist eine lineare, doppelsträngige DNA. Die Konturlänge der gespreiteten DNA ergibt, im Vergleich zu pBR 322 (Molekulargewicht $2,88 \times 10^6$ D), für die TTV1-DNA ein Molekulargewicht von $10,5 \times 10^6$ D. Das entspricht einer Länge von 15,9 kb.

GC-Gehalt

Aus der Bestimmung der Schmelztemperatur der TTV1-DNA von $83,8^\circ\text{C}$ errechnet sich ein GC-Gehalt von $35 \pm 2\%$. Dieser Wert stimmt mit dem GC-Gehalt von 37% der bisher sequenzierten TTV1-DNA gut überein.

Analyse des Genoms mit Restriktionsenzymen

Die DNA des Virus TTV1 wurde durch Spaltung mit verschiedenen Restriktionsenzymen analysiert (Abb. 7). Die Größe der Restriktionsfragmente (Tab. 2) wurde durch Vergleich mit Referenzfragmenten bestimmt. Die spätere Sequenzierung der TTV1-DNA bestätigte die ermittelten Größen der Restriktionsfragmente.

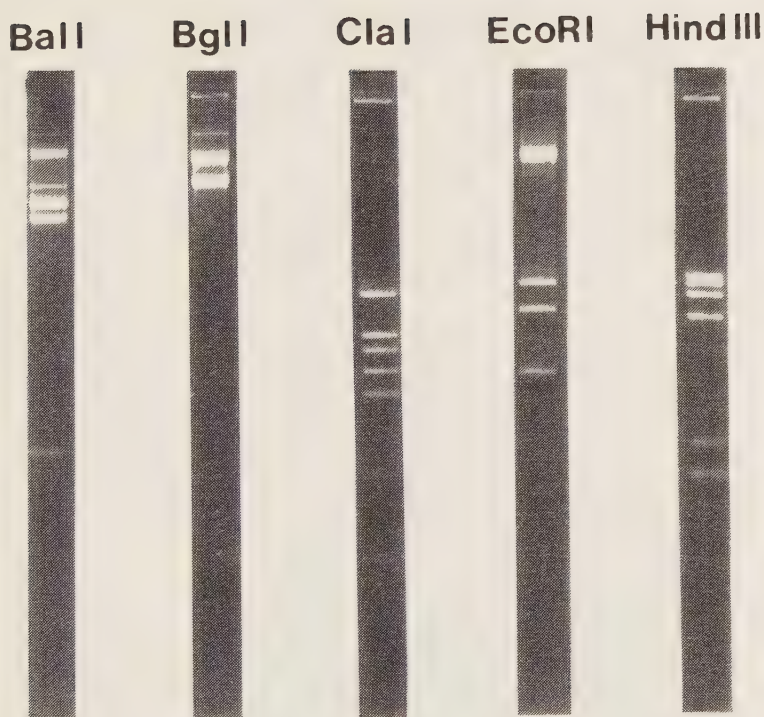


Abb. 7 : Restriktionsspaltmuster von TTV1-DNA, aufgetrennt durch Agarosegelelektrophorese.

Nr.	BalI	BglI	ClaI	EcoRI	HindIII
1	6.200*	9.500*	3.400	9.050*	3.700*
2	4.700	6.400*	2.550*	2.800*	3.550
3	4.250*		2.550	2.300	3.350*
4	750		2.300*	1.500	2.350
5			2.000	250	1.150
6			1.500		850
7			900		850
8			360		100
9			330		

Tab. 2 : Restriktionsfragmente von TTV1-DNA
Größenangaben in bp. Endfragmente sind mit *
gekennzeichnet

Durch radioaktive Markierung der 5'-Enden der TTV1-DNA mit (^{35}P)-Phosphat vor der Spaltung mit Restriktionsenzymen wurden die Endfragmente identifiziert (Abb. 8). Dabei fiel auf, daß jeweils nur eine kleine Fraktion der isolierten DNA von TTV1 mit Hilfe von Polynukleotidkinase terminal phosphoryliert werden konnte. Auch nachdem mögliche terminale Phosphatreste durch Behandlung mit alkalischer Phosphatase entfernt worden waren, änderte sich dieses Verhalten nicht. Ebenso konnte ein mögliches an die Enden der DNA gebundenes Protein nach Pronase und Proteinase K Behandlung ausgeschlossen werden.

Dies legt nun eine kovalente, terminale Maskierung der

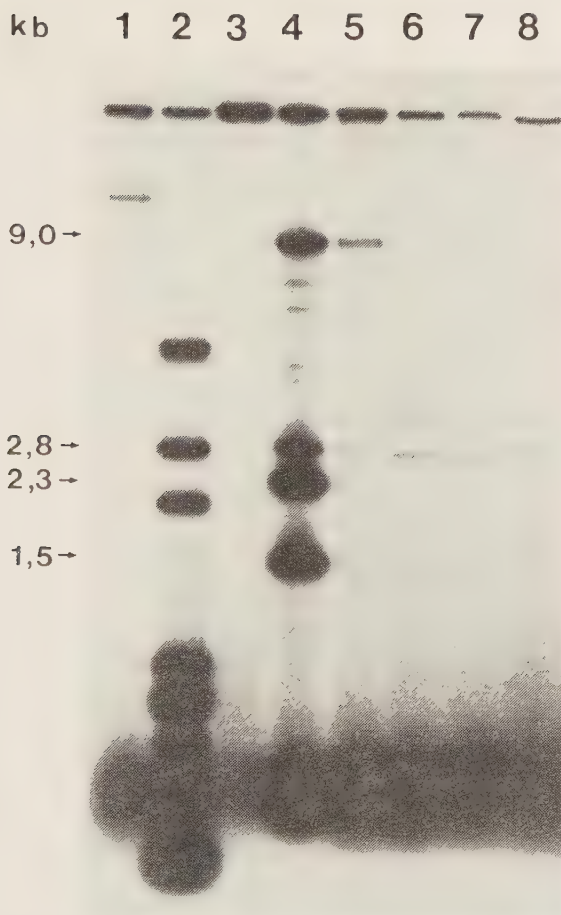


Abb. 8 : Terminale Phosphorylierung der TTV1-DNA
 (1) ungespaltene TTV1-DNA, TTV1-DNA terminal
 markiert nach der Spaltung mit (2) HaeIII, (4) EcoRI,
 TTV1-DNA terminal markiert vor der Spaltung mit
 (3) HaeIII, (5) EcoRI, (6) HindIII, (7) ClaI, (8) HindII.
 (Fragmentgrößen siehe Tab. 2)

DNA durch eine bisher unbekannte Komponente nahe. Weitere Hinweise in diese Richtung ergaben sich durch das hydrophobe Verhalten der DNA bei der Phenolextraktion und der Ethanolprazipitation. Auch war es bisher nicht moglich, die Gesamt-DNA des Virus TTV1 bzw. das eine Endfragment (Fragment 2 im ClaI-Spaltmuster) zu klonieren. Interessanterweise konnte aber das andere Endfragment (Fragment 4 im ClaI-Spaltmuster) kloniert werden, was aber eher damit erklart werden kann, da sich am Ende dieses Fragments und damit am einen Ende der Virus-DNA eine Schnittstelle fur das Restriktionsenzym ClaI befindet.

Die Restriktionskarte

Die Schnittstellen der Enzyme BalI, BglI, ClaI, EcoRI und HindIII wurden kartiert, indem zunachst jedes Fragment nach Auftrennung im Agarose-Gel herausgeschnitten und mit jedem der anderen Enzyme nachgespalten wurde. Wegen der Unsicherheiten, die durch Fragmente gleicher Groe bzw. durch sehr kleine, im Gel nicht mehr sichtbare Fragmente auftreten konnen, war die Verwendung von mindestens drei Restriktionsenzymen erforderlich. Die Restriktionskarte (Abb. 9) wurde durch die spatere Sequenzierung der TTV1-DNA bestatigt.

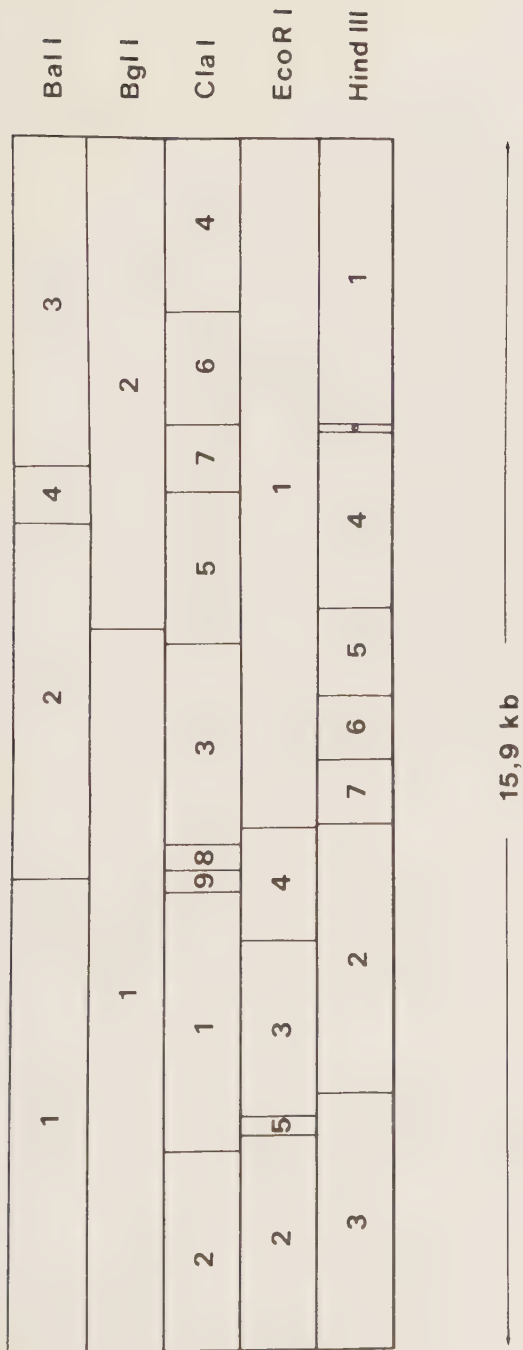


Abb. 9 : Restriktionskarte von TTV-DNA.

Homologie zwischen der DNA des Virus und dem Genom von *Thermoproteus tenax*

Radioaktiv markierte TTV1-DNA hybridisiert mit einem Restriktionsfragment aus der ClaI-geschnittenen DNA von mehreren Klonen von *Thermoproteus tenax* (Abb. 10). Dabei ist die Hybridisierungsintensität zwischen den einzelnen Klonen bei gleicher Menge eingesetzter DNA unterschiedlich. Das läßt entweder auf eine variable Anzahl von Kopien der hybridisierenden DNA im Genom schließen oder aber auf ein ebenfalls in variabler Anzahl vorkommendes extra-chromosomales Element, das diese DNA trägt. Die Homologie der Virus-DNA und der DNA der aus infizierten Kulturen stammenden *Thermoproteus tenax* Klone konnte auf das ClaI-Fragment Nr. 5 der TTV1-DNA zurückgeführt werden, das heißt die anderen ClaI-Fragmente hybridisieren nicht mit der *Thermoproteus*-DNA. *Thermoproteus tenax* Klone, die gegen das ClaI-Fragment Nr. 5 hybridisierende DNA besitzen, zeigen eine Art Resistenz gegen TTV1-Infektion. Zellen dieser Klone konnten zwar reproduzierbar mit TTV1 infiziert werden, aber die Lyse trat erst 14 Tage nach Infektion auf und dann auch nur sehr unvollständig (Zellabnahme im Lysat von 10^8 auf 10^7 Zellen/ml). Der genaue Mechanismus dieser partiellen Resistenz bleibt noch zu untersuchen.

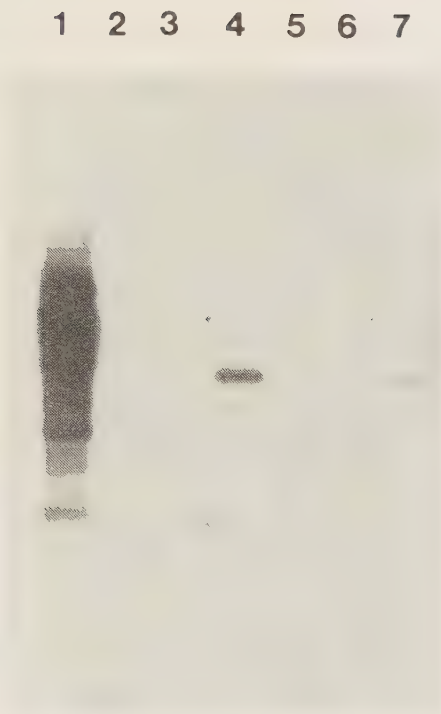


Abb. 10 : Homologie zwischen der DNA des Virus und dem Genom von *Thermoproteus tenax*.

(1) ClaI-Verdau von TTV1-DNA

(2)-(7) ClaI-Verdau von *Thermoproteus tenax* Klonen

(4)-(7) positiv reagierende Klone.

DNA-Sequenzanalyse

Von der Nukleotidsequenz des Genoms wurden 13.307 Basenpaare sequenziert, das entspricht 83% der Gesamtlänge der TTV1-DNA. Die DNA-Sequenz wurde ermittelt mit Hilfe der Dideoxy-Kettenabbruch-Methode von Sanger *et al.* (1970) durch die Sequenzierung beider Stränge in jeweils überlappenden Restriktionsfragmenten (Abb. 11).

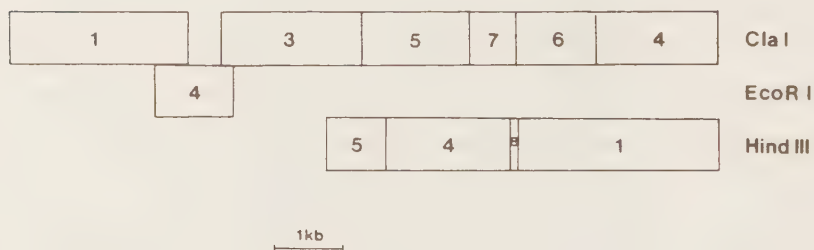


Abb. 11 : Die sequenzierten Restriktionsfragmente der TTV1-DNA

Das Restriktionsfragment ClaI-2 konnte aus bekannten Gründen bis jetzt nicht sequenziert werden. Die ermittelte DNA-Sequenz beginnt daher mit dem 5'-Ende des ClaI-Fragments Nr. 1 (bp Nr. 1 auf der Sequenz, Abb. 12) und endet mit dem 3'-Ende des ClaI-Fragments Nr. 4 (bp 13.307 auf der Sequenz), das mit dem natürlichen Ende der TTV1-DNA aber nicht übereinstimmen muß. Die Basenzusammensetzung der Sequenz ist aus Tabelle 3 ersichtlich. Sie entspricht einem

GC-Gehalt von 37%.

A: 4667	C: 2072	G: 2833	T: 3735
GG: 625	GA: 1078	GT: 674	GC: 456
AG: 1010	AA: 1346	AT: 1542	AC: 769
TG: 853	TA: 1514	TT: 884	TC: 483
CG: 345	CA: 728	CT: 635	CC: 364
GGG: 98	GGA: 292	GGT: 147	GGC: 88
GAG: 249	GAA: 319	GAT: 324	GAC: 186
GTG: 157	GTA: 274	GTT: 184	GTC: 59
GCG: 75	GCA: 163	GCT: 139	GCC: 79
AGG: 220	AGA: 355	AGT: 259	AGC: 176
AAG: 303	AAA: 379	AAT: 480	AAC: 184
ATG: 363	ATA: 672	ATT: 347	ATC: 159
ACG: 124	ACA: 263	ACT: 253	ACC: 129
TGG: 245	TGA: 277	TGT: 202	TGC: 129
TAG: 298	TAA: 399	TAT: 531	TAC: 286
TTG: 206	TTA: 309	TTT: 216	TTC: 153
TCG: 88	TCA: 165	TCT: 127	TCC: 103
CGG: 62	CGA: 154	CGT: 66	CGC: 63
CAG: 160	CAA: 249	CAT: 206	CAC: 113
CTG: 127	CTA: 259	CTT: 137	CTC: 112
CCG: 58	CCA: 137	CCT: 116	CCC: 53

Tab. 3 : Die Basenzusammensetzung der TTV1-DNA

Die gesamte DNA-Sequenz enthält insgesamt 19 "direct" und 2 "inverted repeats" zwischen 15 und 22 bp Länge. Auffallend daran ist, daß zwischen dem Basenpaar 10.497 und dem Basenpaar 11.961 insgesamt 18 der direkten Wiederholungen kurzer DNA-Sequenzen vorkommen.

Abb. 12 : Die vorliegende Sequenz der TTV1-DNA

```

1  ATCGATATCTTCATTCAACTAACTGAAATTGCTGAGATGGAAAGAGACGT
51  TATCGTGGAGGAACAATTGAAAGATGAAGCAAAACAATTGCAAGAAATATG
101 AAGATTTCTGGTTTGACGTATACACATGACGAGATACAATTACATTAGGA
151 ATCTAAGAAATAATCAGAGGAGAGGAAAATCAAAGTAGGTGTGAGTAGCGAT
201 AGAATCCAATGATAGAGATTGAGCCAGGACACAATGACCCAGAAAGAAT
251 GTGTGTAATGTTTACGGCTATGGAATATAGTCATATACCATATCAAGCAG
301 ATAATCTAGACGAAGTATATGACTCATGTAAAATTTTGACAAGCAAGATT
351 ACTACTTACATAAATGTAGCAATAGGATTTGGCAAGAATTTGGGGCTCAA
401 ATTAAGCGAGATGATGTTGGTGACTAGAGATAAGTCGAAAAGAGAATTCA
451 AAGTTGGATTAGAATTGGGAATAAAATATGCATATAATTGGACCTGTAAG
501 TTTTCAGTAGGACGGTATTCCCTAGATTTCTACAATTATTATTACAAAGTG
551 GTCTTTTATGAAACGCTAAGAAATAGATAGAAAGTACATTGAATATACATTG
601 AAAAAATAAATATGCTACACTAAGATTATATGGCAAGTACGTTTACACATG
651 ACGTAATAGTAATGGAATGTTATGAGGAGGCATTAGCTGTGAATGAGGTGT
701 GAGGTGGCATATGACAGAGTGTGACTAAATGGATGGAACCAAGCTTACC
751 GTTGTGAGACTTGATAATGAAATTATAAAGCCTTATATGAAAGTAAAAAT
801 ACGTGATAAAACAATTATGATGAGCAGAATAGACAAAAGAAGTAAAGAAAA
851 TAATGCAAAATATATAGAGGCACTGAAGGCTCAGTCAATGATAATGACCTT
901 AGAATGTTGGCTTAAAGAAGGCGGTAATTATGGCAGAGCTGAATAAGGCTA
951 TCGACAAGCTGACATATAAAATTTGGCAATGAATTACCATATACGGCAGAC
1001 ATTGCTGAAAAGCTGATAGAATTAGGATTAATTAGACAGGAGAAGTGGCT
1051 GAAGGCAATTAATGACACAAAGAGGTTTAAAGATAAGATATTGGACAAGC
1101 GTAATTCTAACAATAAAATAGAAAAATCAGAGAATACAGGAGAGAAATCAT
1151 GATAATAATCAAGACCAAAAAATAGAGAATATACGATAGATGAGAAATAAGG
1201 AGATAAATTCATAAGTGACGCTGATACATATATGTCATATCGTGATATAT
1251 CGGTGCTATCATAGAAAAATCCCAATAAAAAATCGACGATTGTAAAGAGGC
1301 GAGAAAGATAGCTGGGAGCATGACGCATTACGTATATTGCGCAGAGCTAA
1351 AGGACGAGGATGTGAGACGGAAAAAGATAGTTGACCCAGAGCTGAAGTTT
1401 AGAGCTGATGCTATATGTGATGGGGACGCATTGTTGAGATTAAAGCAAAAC
1451 TTCAATCGGGACTGGCATGCGCTTAATAAGTGGCAATTGAAAGTGTATAT
1501 AGCGCTGGCAGGTGCTGAGAGAGGACACCTAATATACATTAACAATTGGCA
1551 TAAGGACGACAATAACAAATGGACGAAGAGGAGCGAGTAAAAATTACTGAG
1601 TATGTTAGAAATCTGATAAGACGAATAAAAAATGGAGAAGCTGTGGAGGA
1651 GAGGAATAGAGGAATATGGTGTAACATATCGCAATTTAAAGCTGAATGTA
1701 ATCATAAATTGCTATGAAGGCGAAAGAAGTCATAGAGATACTGGAAATCA
1751 TAATATACATTGTAGTAGTGCCAAAGCCTTATAGTTTATGTTTACAT
1801 AATACGGTATATGTAGCAATAACGGTAACATTAACATTGAATTTAGCAAG
1851 GAAGACTTATAGTACAAAGAAAAAGATTGAAAAAATGATGAAGAAAGTAG
1901 AAAAGGCGCTCGAGTTTTTCTCATGAGGTATACACAATTATGGGATGACA
1951 GAATAGTATACCTAGGAGAATATGAGCTAGTGCCAATGCCCAAGCTGGAA
2001 GACGATATTGCGGCCATTAAAGGAAGCATAAATAGCGTATGTGTTAAATAG
2051 TAATTATTCAAGTTTCATTGGTGCACTGAAGGATACAAAAGAAATGGAGT
2101 TAATTGAAAAATTTAGCATATGAATTACACGAATATAGACAGGACACCG
2151 ACACCTCACACCATACTATTATTACTCATGAAAGTCTATGGGATGCATGG
2201 TTCAATCTTGTCTCAATTTTAACTTACATAAGTTTCGAAATAGAAGATAT
2251 AAACAATTTAATTTATGGAATTTGGAAAGACGACCGTAATGCAGAAGCTCA
2301 TAGGAAAGTAAATGAAAAAGTTTATAAATAGATTGTCTTATGCTTTATG
2351 TCAGTAGTTGTTGGAAGAATTGCTGGAGCTAATGAGTTGGAATGCGGTA
2401 CCAAGCTGGGATTGAGGTTAAACGTGATGTTCAAGTCAAAAGAGTTGAGA

```

2451 GAAAGATATCGTATAGAGCGCCTAGAATAGAGGACGCAATAGCGTTGAAC
 2501 ATATATGTTGAGGATAAGATTAGTGGAATTAAAGCCGTATATACCTGGTCTA
 2551 AGGGATTGGTAAGAAAGGCGATTAAATAGCATATCGTGATGGGGTGAC
 2601 AGTGCAATGAATATATTTAAGTCATATGATACCGCAATACAGAATGAAAT
 2651 AAGGAAAATAATCGAAGAGGGTCTAAAAATGTATGAAAAATAAGGAGTG
 2701 AGGGTAGGCCATGAGCTCCCCATAAATTAGTAATTATCTTGAGGGAATTC
 2751 GCGATAAATAATCTGCAACGGATAGAATGGGAAATACAGGTAGCTCAG
 2801 AGCGGATTCAGAGCTATTCCAATTACTGAGAGATGGTGGGGTGAGGCATC
 2851 ATATTATATAGACCTAGCCAACAAGATAAAGGCTTTGCCCGAATTGAGCA
 2901 ATGTTAAGACTAAGCCGATTGGATATGCGTTGGCCCGCGGAGATACGTT
 2951 AAACACAACAACCTTCGGCATGGCCGAGATATACCTTGATAGATTTGAGTC
 3001 TCGCCTTGTCGCTAGCGGACTGTCGAGAGATATGGCGAGAAGGGTTAGAG
 3051 AGCATGTGGCAAAAATGATAGGGCTTGTTGGTGGAGGCGGTGGTGGAGGA
 3101 GGCGGTGGAGGCGGTGGTGGAGTAAGTTTGCTTAAATAGATAGAGTATGA
 3151 AGTATAAATTCCGCAATAGGTTTTTTTTATTAGTCCTTCGGAGTTTTCTATTA
 3201 TATAATGACTATAAGAAGATAAGGACTATATACAATACAGCTTAGAGTAGG
 3251 AGGAATGAGATAGAGAAGTGAGAGTAGGTAGGCACCTTTATGAGCGGGTG
 3301 AATATAGAATATACATAGAAGGTATACCGATAACAGTCTATAGAGGAGAT
 3351 AAGACAAGTGCAAGCATATATATGAGCTGGGGTAACTGGATATCGATGTA
 3401 TAATTTATCTTTAGAAAAGATTCTGTGGGCATATAATGCAAAAGCGATAG
 3451 TTATTAGACCAACAGGATTCAAAGAGGGAGAAGGATGCCTTAGACGAGTA
 3501 TTAGATGAGATAGATGCTGGAGTATATTCTGGTGGGAGATTATTGAAACC
 3551 TACAATTGTACAATTGAGAGAAGAGTGAATACATTGTTTGACCCAGTTGA
 3601 AGTAGACATAATGGCCCGAGTATATGGCCATTTATATAGTAGCAAAAAGG
 3651 ATAAAGAGTTGATGAGTGCGTTAATGCTATGAAAGTACTTAGGTTAGGAT
 3701 TCACCGATAAATATATCGAGAAGCAGGAGCCTATCGATGTAGTATTCGAT
 3751 AAGTACCAGCCGCTTGGATATGTAGCTGTAGTTGGAATTACCTTCGAGGAT
 3801 ACCGTGGATAATCGAGATTCCAGAGGGAATGGAATAGAAAGGTTACATAA
 3851 CAATGCCGCGTGACATATTTAGAGAATTATCATTTGACATAAATTACTG
 3901 AGGAGAAAAGCTGGAGCCTACTCCTCAATATCGTATGATAAAGGATATAGT
 3951 GTCAGATTTAAGACAGTCAAGAGGATATGCGAGCGGATGGTGATCTGCG
 4001 CTAATGGACTAACATATGATGGGGACCTGCTTGAAGGTATAGAAGTTATG
 4051 GAAGGAGTTGATGTAATAGCCTATACGTTGGGTCTAATCGATTTCTAACA
 4101 GATAAGTATATTAAATACCTAATCATTTTCTCTCATGATTACAAGTAACAT
 4151 TACAACATAATGCTATAAAATCAAGTAGACCCGAAAGCGGTACAAGAGGGT
 4201 TAAAGATTTTTAGAACTACTGAAAAAGATAAGGGAAGGACAAGTTGATGTT
 4251 AATGAAATAGATAGTTTACCATTGTGCTAAAGAAATCTATATAAAAACGGG
 4301 ACGATCATTAGTTAAGCTTATACAGGATATGATGGCGAGTGACCCAGATA
 4351 AGGCACTCGAGTTATTTTTAGGAGCATTTATGCGTGGAGAGGAATTGAAT
 4401 AAGGCGATTCAATTCCATAAGGAGCTAATAGCTGAGTTATCAAAAGAAGA
 4451 TGGGGCGGATTTATGTTACTAAGGTAATAGAGTATTAGAGAAGTACGGCT
 4501 CGAGAATAGATTGTGAAAATGCAGAAATGTCGGTAGAAATGGCGGAAGA
 4551 CTTGTTAGGGAAATAATTGAATGATGAAGGTTAAATTAAGGAAAAGATA
 4601 TACTTTTATACCGCCATTTAGGGAGAAAGATTTAAATACCATTGGATAGTAA
 4651 TTAGACAATGGTTGAGATTAAACTAGTTAATAAGGAAATAATAAAGTTT
 4701 GGATTGGCACTTGGAATTGTGAATGAGCTTAATGAATACCTATACGCGGC
 4751 AATGCCTCCAATGTATGACATATTATCTAAATTTGTACGGATATGGAAGGA
 4801 CTGTAAATGCAATGCTGTATAGTGTGCTGTTTAATTTACTCGAATCGAGG
 4851 ATTGATACATCGTGAAGACTTGACCTTACTAAGTTCTTTGCTGTGGTAGC
 4901 ATATATGGATGCTGTTAAAGATACCGTTGAGCTTGCAATAAATGGATATG
 4951 CATATGTTTACGACTTCTGGATTTACAGTGATATCCTGCATCAAGATTACA
 5001 GGAACGTATGAGGCGACTTCAGTCAAGCTAAGCGATACCCGCAAAATAC
 5051 TACAGCAACACAGGTGTGGTTTGACCCCAAGAAGTTTGCTTTGATTACAC
 5101 AGAATAAGATATATACATTCTTAGCTCCTTATTGGTTCTAAACTTAACTA
 5151 AAGCTTATTAGCAATTCAATACTTTTAGATATGAACATAGTTAGACATAG
 5201 GGGTATTGTGTATTGTGATGCAGACTTTTATCATATCATAGAGCGGAGG

5251 CTATAAAGGATGTTGTAGGTTAGATATCCAGTATTATGTTGTTATAATGAT
 5301 GTATGTTTATGACCCGTAAAGGGAAGGAGTATACGATGAATCGAATATACA
 5351 TGAGCTTAGAAGCTATAAAGTTGAACGACTTAGATATAAATAGATATCCCTA
 5401 TAACTAGAGGACTATTATTTTAAACAAATCAAATAGTAAACGCATGGATG
 5451 TAGCAGTAGTTATATCTGTATAGCAGAGATGATGCATTAGACTTACTATAT
 5501 AGAAGCTTTACATTTATTGGTGGGAAGAGAAGTGATGGGCTACGGGGACAGAT
 5551 ATACGTGGTGTGCAATACAAATGTACTACTAGTATAGTAAAGCTAAACCTAA
 5601 TACAGCAGAAGAGGAAAGGTGTAGGGGGAGCTAGATGTGATGCAATAGAA
 5651 TCGACGCTCTGAGGATACATAATATTCTCAGATGGACATGTAACACCACC
 5701 TCGAGAGATATCGAAGATGTTAGTTGAACCGTGGTCAAGTGTGCCCTATAA
 5751 ATCATATAATATTACCATTGTTGGTACATCGGGGAAGAGTAGCTTATTCATTA
 5801 CCATTTTATGCCAGATGAAAACAAATTTTTATGGTGTTCATGTAGTAAATA
 5851 TAAAGAAGAGATAACGACAACAGGAGAAGCCATTGTAGCTATGAGTAGGA
 5901 AGTACATATCTATGAGCAAAATGCTATACAAAGCTATGGACTTGACTTGT
 5951 CAATTTACGTTGAGAATGCCTCCAGGACAGTTAATCGGCGATGAGGGGAT
 6001 TCATTTTATAGAGGCAAAAAGCTTATAAGCAATAGGAAGCCAAATAAAG
 6051 ATGAGATGAAGGACTTCTACGACACGATATACATAAAAAGGAAAAGTGAA
 6101 ATAGAGGAAGAGATAAAGGAGTGCTTCTGTAGGGCACTATGATGCTGTTT
 6151 TTTACTAATATGATAGTTGACGGAAATTACACAGTATACGAGGATGCAAAA
 6201 TTTAGCAATATATATACCTCCAGGTTCCAGATTTTGATTATATCTTTACCA
 6251 AAAATGGGCAATTTATTCTACAGATAATTGCTCAAGGTATTTGGTGTAAATA
 6301 TTTCCAGCAACAGTAAATTATACTGGGCCAGGATTATATAAATTCAGGT
 6351 AAGTAGCGGTGGAGTAATAACTCAAATAACGCCAATAACCAATAGATATAA
 6401 TAATAGGGAATAATATAGTAATATATGAGTTTCAAGGAATTATAATACCT
 6451 GGAGATCAAGCGAGCACAATGGGACTTACAGTTAAGTAAAGTATCATATCC
 6501 AGGACAAGGACTATTATTGTACAAGGATATCCATGGTATATGTACAATC
 6551 CTAATTTACCGATCGTATTTGCGCCGATTGGATATACCTGGATTGTTTATG
 6601 GGTATCGATAACGTTGTAAATGGAGAGACGCAAGCTCAGGCCAACTTAATG
 6651 GTTTTTTACAGCCAGACGGAAAAGTGCCATATGCGTATCAATTGACCATAA
 6701 CGATACAGAATGCTAATGCACAGCAGATAATACAGGCCTTCAATCAAGCT
 6751 ATTGAGTTTGCAGCTGGGATGGAGCTGTTGCGTATGAAATTACATCAAA
 6801 TAATCAAATTACATTCTACATAATCTTCATATCGCCAACTCTTGGTGGG
 6851 TAGTAGCTTTAGTATTGTAGTTGCATTAGCCCTTTTGGCCGCGGCTGTT
 6901 TTTGTGGATAAGTTAACGAAAGGAGTAGCAATGGACATGCGAGGCGCAGGC
 6951 GATTCAAATGCATATAATGCAAAATACGCAAGCGTTTCCAGCGGTATTT
 7001 CCAGCTGTTATCAGCAGACGCTCTGGAAGGACGCAATGCATAAAACAATTGT
 7051 TATGCTAATATACGTGAAAAGTTTTAACGATGTTGGGAGCGCAGTTTGGTTC
 7101 ACTAAATCAAGTTCTAAGTCAGCCAACTCTAAGCTTTAGTGGACAGACGA
 7151 ATGTTAGCGTGCAAACTAAGTGGGCAACGGTAGCGCGTTTACGTTGATAA
 7201 TAGGTGGTGGCGGGGGTAGCGTGGGGTATCTATGAGTGGAAGAAGAAGAAA
 7251 AAGAGTAAACACAGTAAAGCATAAATACCTATCTATTGAATAAAGTGT
 7301 GATAGTAGGTATAACTGATAAAATATGTTATTTTCAATCAAAACCATTTTCAG
 7351 CAACTAAGAAGTTTCTAAGATATAAAGATAAAATACATGTATATACCTGGC
 7401 CCAATTTACTTTCTATCTTCAAACATAGATTGAACTCACTTAAGTACTTT
 7451 AATGTCGAGATGTAGGATATTGTAGTCCACCAGGTAAGAACAATAGTGACT
 7501 ATCTTACAGCATTTCCATTGTGTATCGAGGGAGATGCAATATAGTGAGT
 7551 AATATACAATTATCGGACGGGACGAAAGTGAATGTAATAAAGACGCCATA
 7601 TATGAAACACTGTATGAATATTTTATGTAGATTACTATTTCGAGATTGGAC
 7651 TGAATAGGTTATTGTTTCCAATGTGTATAGAAGTGAAGATTTAGCATGGATA
 7701 AGTGCTGGGAATAATATTGGACCTACTAAAGGAGTATTATTGTCAGGTTT
 7751 TCTAGGTTTCGATATATGGTAGTATGCAATATTTAGTGGAGATTTAATAG
 7801 TAGGAAAGAAATACAAGGTGATATCTAGAGATTTCTACATAGGAAAGACA
 7851 ATAGAGTGGGAAACGACTATAAATGGGAAAGGAATTTTATGTTCTTTA
 7901 TGAAAATGGAAAAGGAATTTGGATAAGAGGATATTACAGCTTTCCAACGC
 7951 AGGTTAGGATAAAACCTGGATTCTCTGGAAGCCAGTGATTGAGGAGTAA
 8001 AGTTTTTATAGTGAAGTACGGAGTATGACAAATGAATAATGTCTCGAGAAT

8051 AATACTTACTGGGATTATGATTTGGCTATTTGTGTAGCAGCTATCTAGTT
 8101 GTACCTACTAGTCCACTTAAAGTGGCATTTATATATGGGTGTAGAAGTTTT
 8151 AGGAGGCATATTCAGCAAAAAGGTTTACGTCCTATAGTGGGTTTCATTTATTT
 8201 GGTTCATTATAACACCTATTCTACTATATCTGCCATGGCCCTAGCTATG
 8251 CTGGGATTTCTAACGCTCTATGGTGCAATTACCTTTTGGTGTGCGGATAGT
 8301 CAATTGGCCATATCCTCGAGATGCCGCGTTGTAAAAAATAGACTATT
 8351 AATATAGAGTGAATTAACATCGCATGATTTTTCATAGGACTCGAGAAACA
 8401 ATAGACTTGGATTATGGTCATATTTATTTATGGGTATTGAAAGGAAATCTCG
 8451 AGGAAATTTATGACCTATATATACTGATTAAGAAGGAAGGAAATTTGAAG
 8501 TTGACAGTATATTTACTGGAGATGTAATGGTGGTTGATAGAGGAAGTTTT
 8551 GTATCGACTTCTAGAGGATTCCTGATGCTTCTTGTAGAGGTTCCCGATCG
 8601 ATAGATTTTAAATGCGGTTAAACCTAAATGGACATGATTTTGGGTATTAGT
 8651 GCATATAGAAAAGAAACAAAAACGGGTCTGAGAAACACGATTTTATACGGC
 8701 AATAAATACGGGAAAAAGTAAAGAAAGATACCTTGGGACTCGATAAACCTG
 8751 AAGATAAGGAGATATTTCAACGCATAATAAACGAATTTAAAGACTATATT
 8801 ATAGACGTAGTATACGCAGAAGGATTGGGAGCTTAATAGAATGAACATTT
 8851 AGTACGTGTTCTGGAAGAAAATGAAGACTACATATGCTTGGCTGACAGCC
 8901 ACCATACTTTCATTACACCGAGTGGCAGACATAGTTAAAACGATGTGTAAAC
 8951 TCTCAGGCATGTGATTTCAAACGATTTGACTTTGGATTTGATGTTGAGAA
 9001 CAACCCACTATCAGTACTTAATAGATCGGTACATCACTTCGGTCAGTTTA
 9051 ATATCGACAGGTTGGCCAAATAACTATATCTTCAGTTGTGAATACAAATAC
 9101 TATGCCAATAATCCATTCTTATGCTTCAAGAAAAGCGCAATTAAGGAGCT
 9151 ACTTAATATATATGGAGGAAAAATAATACCGTGGCAACATCTGGAGCTG
 9201 ATATGGAACAGATATACGTTTCAATAGTCAGATGATGGAGAACAGCT
 9251 GGCAAAATGTATCGGAACCTTCAACCAATATACGGACATAGAGCTACTGTAAG
 9301 TAATACTGAACATGAATTTTGGAAATCGAGATGGACGATTAAGATTATA
 9351 GCGAAGGATGGTATCAAGCAAATGCGTGCTTCTTGCATCTCATATACCT
 9401 CGAGAGTTTGTGAATCTGAAACCAAGAATATTTGAACACAGACAAATGTAA
 9451 ATTACATATAAACTTTATAAAATCGACTCTAGAGGATTCCCAAGCTTCTT
 9501 CTAGAGGTACCGCATCGATGATATATGTCAGTTGATGGAATAGATAGAAAC
 9551 ATACTTAACATCTAGTAAATATACGGCTGATAGTACTGGAAGCTGTAAG
 9601 CTTAACCAATCTTCAACCTGGTAAATATATAGCTATTGAGTAATGTGTCG
 9651 TACCAAAATACTATAAAATAATAACTACAAGTAACGTTCAAAATATAGT
 9701 TATAGCTGATCCTAATCTAGCATCTTTAGTGTAAAAATGTAGTATTTAGTG
 9751 ATTTCTCAAGGACCATTATCATCGTTGTATGTATATGATGGGTATGGGTA
 9801 GTTAAATTAAGGTCACCGCTTTCACCTGGACAGAGTATAACAATAACAGC
 9851 ATCTAGTGGTACACCTAATATAGACCCCAATAGCAGTATCTACTATAATA
 9901 ATGGAAGCTCATATTCAAACCTTACACTAGTAGGTTCACTACAGTATCT
 9951 ATTGTACAGGACTTTGGTGGATATGCAATAAGTGCATATGCATCTGGGGA
 10001 CTTCTTCTTAGTAGCATCGCCTACTGGATTTACTCCAAGTAGCAGTAGAC
 10051 TACTAGTTGTTATAGATGGGCAACTACTCCTACGAGTCTTGACGCGTA
 10101 GGTTTAGGTTTATATGCAGATACTAACGACTGTTTGGTGTAGTTAGGAA
 10151 ATATATTTAATGGAGCTCAGAATGTAAGCATTTGAACAAAAGATATCTGGAA
 10201 CATATTCAGTAGTAAATGAAATAGATATACAGGTTTGGCGCATTTTACT
 10251 GATCCATTAGTAATGTACTTATCAATAAATGGAAGTACTGCAACGTA
 10301 GGTATACAAACAGGGCTCAAAACATTGGAACCTGTAAGTGGTAACCTATTCTA
 10351 CTACTCCATATGGAAATCCATCTATGGCAGGATATGGAACAGTTGATAAG
 10401 CATTATGCTAACTTTATTTGTGTTTACCATATGAGCCAGACCCACAGGTAAC
 10451 AGTTACTCCTATATCTTCAACCATCTCCTACTCTTACACCCAGCCAGCAGC
 10501 CTACGCCCTACACCGACATATGATATCACTTATGTAGTATTTGATGTAACG
 10551 CCTTCACCTACTCCAACCTCCTACTCTTACATCGACACCCACTCCTACTCC
 10601 CCAGCCAACATATGATATAACTTATGTAATATTTCGATGTAAACCTTCAC
 10651 TCACACCTACTCCAACACCTACTCTTACACCAACTCCGACGCCAACCTCT
 10701 ACATCGACAACGTCATCTAACATCTAGTTGTATCAATATAACAACAG
 10751 GTATTCAATTCCTTGACAGTATAGTATTTGTGTATAAATGGATTGGGAATA
 10801 ACAGTGTTTATAATAATGCTAATTGTAATAGCAATAATAATTATAGCATA

10851 TATAATATGATAAACATAGATGGAGACTCAGTGTGAATATGGAATACAGT
 10901 GTACTAATGAAGAAAAATGTATATGGCTCGGAGGTACATGTATCAAGAAG
 10951 TGTGAACATGGATGTTGTTGTATATTTATTACTTTGAATGATACACACTA
 11001 CTAAATAATAACGTTACTGATTATATCACTGAACCTTATTAATCGATCCT
 11051 CTAGAGTCGATAACTTCGAAAAGCATGAGCTTTACTGATATATAAGGCTG
 11101 ATAGGACTATTAGAGTTTCTAAAATTAGTTGAATCTGCCGAATTAATAAT
 11151 GATGGGTAGCTATGATAGCTGAGCAGATATTTATAAATATTATTATTA
 11201 ACAGTACTACACACGGCGATTGAAAGGATAGCTAAGCGTAGTCTATCATA
 11251 TAGTAATGTAAATTACACATATACTCAGTCGATATCGTTCTTAGTGGGAT
 11301 ATACTATAGTTTACATAAATTATGTCACATATAAGCCATGCGGTCATTATA
 11351 TCTATAATGATTATGATATCATATGCCATATATATTGTAATTAAGATGAG
 11401 AAATAATCAGCAAAATACAGAACGATATTGAGATGTTTATGTTTGGAAATGA
 11451 TAGCTGGACTAGGTATAGAGATAGTAGGATAAAGTATATAAATAAGTAGG
 11501 AATTAAATTAGCGTATCTGCAACTCCAACGCCCTACGCCGTACACCCACACCA
 11551 ACGCCAACTCCGACGCCTACTCCTACCCCTACTCCGACACCTACGCCGAC
 11601 TCCCCTCCAACACCTACTCCGACTCCGACACCTACTCCAACACCAACCGC
 11651 CAACTCCGACACGCTCCTACCAACGCCCAACGCCCTACTCCTACACCAACCCCA
 11701 ACACCTACCCCTACTCCTACACCTACGCCGACTCCGACTCCGACACCCAC
 11751 TCCTACGCCTACTCCGACTCCTACCCCAACTCCTACTCCCACTCCCACCC
 11801 CTACTCCGACTCCAACACCGACTCCGACACCTACTCCGACGCCCTACTCCG
 11851 ACTCCAACGCCGACTCCTACACCAACGCCCTCACCTACACCAACCCCA
 11901 CCCTACTCCGACTCCGACGCCAACTCCGACTCCCCTCCAACGCCCTACAC
 11951 CAACTCCGACGCTATGTAAACCGATAATTACTACTGGAGTATCATGGTTA
 12001 GATACAAACGATAAATGTGTGCTGGAAGTGTTGGTATTCGCTATTTGAAAT
 12051 ATTACTATCAATACTAGTAATACTGCTTATAATAGCACTGATATAACAGA
 12101 ATGATATAGAGATGATAAATATTGGAATGATAGGTGACCTAGGGATGACG
 12151 ATATAAAATAAAGCAAAAAATTTTAAAGACCTAAAATACTCTATTTGTGTC
 12201 CATCAATTAGTACACAGTCAATAAAGTACGCCGTCTGTAGGTGTAGCAATT
 12251 GGACCCGATGGAATATATAGAGCTCTTTATGGATGGTATGGAATGAACA
 12301 GCCGCTGGAAATAGATAAATGTTTCATATGATTTATCGACTTGTTCTCCTG
 12351 TACTATCGAGTCGGACCAACAATAGGTGATGGATGAGGATTTTATGAT
 12401 GGAACATAATATATGGTTCAGTGGTTCAGATGAGGCGAATGATCAAGGAGT
 12451 GGTAGCAAGCTATAATCCATCTACAGGAGCATTCAAATTATGCATATTATC
 12501 CTGGTAATTCTACAAAATACGTAATTAAGATATTCTACTATAATGGATAC
 12551 TACTATTTGATAACATGCTGTGACCCAGGAACACTACTTAAGTGCACAAA
 12601 TCCTCTAAATCCGTCTACATGTACCCAGCTTAATATAAATGCACCTACTT
 12651 CATTTACAGTTCAAGAATACTATATGTATACACAAGGCTCATATGCACTA
 12701 CTATCACTACTCCAGAAAAATATGAATAATAAATTTGTTCACTTTATGTAA
 12751 TTTTCGATGGAACAAATCTATATAAAGTATATAAATATATACTCTACGACTG
 12801 GAATACTACATTATCTCCATCAGTCTTACTAGAGGAGCAATATGTGGAG
 12851 ATAATATAGTATTTCCAGATATTATAAACAAATGGAAGTAGTGGTTGAGTTC
 12901 GATGTATAGTAGCGCCTCTTTCAAATCCAAGTCTTACTACCAAGTTTATG
 12951 TAGTTCAACATATACACTATCTACTACTACTGGTGAGCTTCATGCGAATA
 13001 TAGCATGCTTTGGAGGAAGGTTTGCTATTGCGCAACTATCTCGATCTTCG
 13051 GTACAAGGAGAAATAACTCTGATAGACCTTAAATAAATTCGGTATATATGC
 13101 TAAATAGAGTACATATATCATTTCCATCCAGATAGCTGAATATTCTATG
 13151 GTGGAAGTCTAATAGGTTTCATGTCAGAATGCGAGTGGAGCAAAATCAGTCT
 13201 CTATTACCAATAATATCGTTTAACTATAGGCTTGCTTACGCTAGTGTGTC
 13251 TGGAAAGCTCTGCGACGTTGACTGGATTTGACCCTAATGCTATTGTAAGAC
 13301 TATCGAT

Offene Leserahmen auf der TTV1-DNA

Die bisher sequenzierte TTV1-DNA enthält 38 offene Leserahmen in allen sechs Leserastern auf beiden Strängen, davon 28 offene Leserahmen auf dem einen Strang und 10 auf dem Gegenstrang (Abb. 13). Die untere Grenze der offenen Leserahmen von 150 bp entsprechend einem Protein von 50 Aminosäuren wurde willkürlich gewählt. Der längste Leserahmen, beginnend beim Nukleotid 6.138 und endend bei 7.193, könnte für ein Protein von 352 Aminosäuren Länge kodieren (Tab. 4). In den 38 Leserahmen gibt es 8 überlappende Leserahmen, das heißt, das Startkodon des nächsten Leserahmens beginnt innerhalb des vorhergehenden. Die Startkodons der Leserahmen 12, 15 und 25 beginnen jeweils ein Nukleotid vor dem Stopkodon des vorhergehenden Leserahmens. Der größte Abstand zweier Leserahmen beträgt auf dem einen Strang 1.195 bp zwischen Leserahmen 37 und 38 und auf dem Gegenstrang 4.066 bp zwischen Leserahmen 26 und 15. Auffallend ist, daß 74% aller offenen Leserahmen auf einem Strang liegen und damit in die gleiche Richtung gehen. Die Dichte der Leserahmen nimmt außerdem gegen Ende der Sequenz in den ClaI-Fragmenten Nr. 6 und Nr. 4 deutlich ab. Der Sequenzvergleich aller hypothetischen Proteine mit der NBRF-Datenbank ergab keine auffallenden Homologien mit bekannten Proteinen.

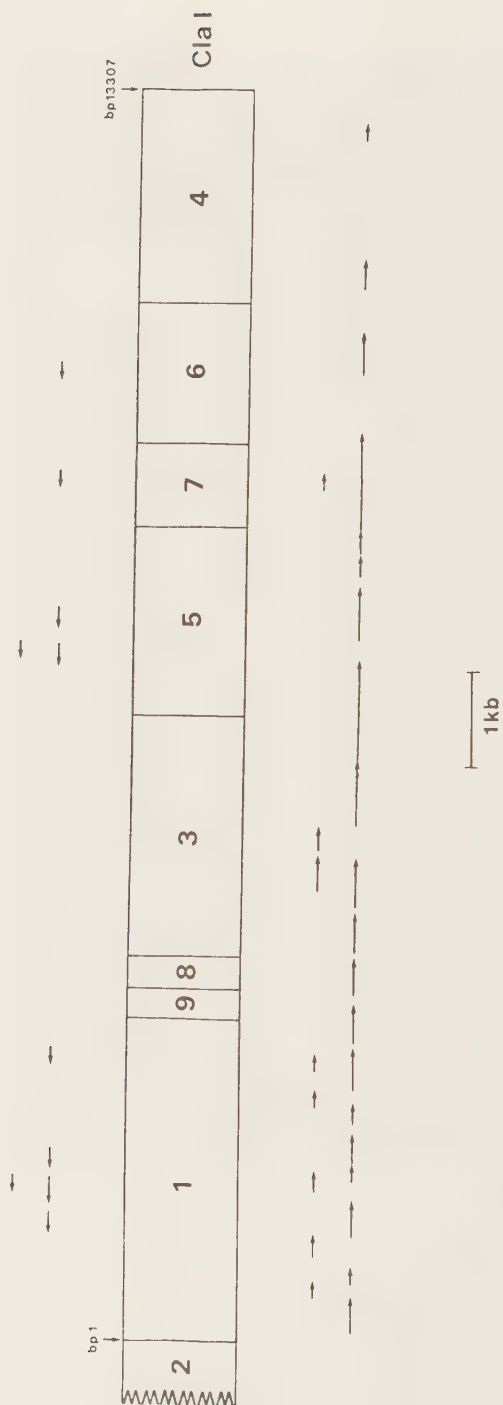


Abb. 13 : Offene Leserahmen auf der sequenzierten TTV1-DNA.

Nummer des Leserahmens	Region auf der DNA-Sequenz	Länge der Amino- säuresequenz
---------------------------	-------------------------------	----------------------------------

1	126 - 506	127
2	478 - 648	57
3	629 - 817	63
4	929 - 1171	81
5	1328 - 1110(C)	73
6	1149 - 1487	113
7	1677 - 1414(C)	88
8	1705 - 1532(C)	58
9	1632 - 1799	56
10	1714 - 1923	70
11	2029 - 1781(C)	83
12	1923 - 2219	99
13	2312 - 2548	79
14	2505 - 2711	69
15	2711 - 3133	141
16	3055 - 2867(C)	63
17	2901 - 3065	55
18	3147 - 3575	143
19	3679 - 4095	139
20	4134 - 4571	146
21	4659 - 5138	160
22	4942 - 5091	50
23	5181 - 5450	90
24	5444 - 6139	232
25	6139 - 7194	352
26	7334 - 7122(C)	71
27	7344 - 7192(C)	51
28	7452 - 7997	182
29	7695 - 7507(C)	63
30	8134 - 8334	67
31	8374 - 8601	76
32	8627 - 9640	338
33	8991 - 9173	61
34	9552 - 9394(C)	53
35	10324 - 10169(C)	52
36	10263 - 10724	154
37	11162 - 11479	106
38	12674 - 12856	61

Tab. 4 : Offene Leserahmen auf der TTV1-DNA und die daraus resultierenden hypothetischen Proteine. (C) markiert offene Leserahmen auf dem Gegenstrang.

ANALYSE DER STRUKTURPROTEINE TP1, TP2 UND TP3 DES VIRUS TTV1

Kartierung der Gene für TP1 und TP2

Die viralen Strukturproteine TP1 und TP2 wurden mit Hilfe des Expressionsvektors pIN III (Abb. 14) kloniert und auf dem TTV1-Genom kartiert.

Dazu wurden die EcoRI- und HindIII-Restriktionsfragmente "shotgun" in jedes Leseraster (Abb. 15) des Expressionsvektors kloniert, transformiert und die erhaltenen Bakterienkolonien mit polyklonalen Antikörpern gegen die Virusproteine TP1 und TP2 getestet. Die positiven Klone wurden lysiert, die Proteine auf einem SDS-Polyacrylamid-Gel aufgetrennt (Abb. 16) und im "Western-Blot" Verfahren erneut gegen die Antikörper getestet (Abb. 17). Auf diese Weise konnte das Gen für das Protein TP1 im EcoRI-Fragment Nr. 3 und das Gen für das Protein TP2 im EcoRI-Fragment Nr. 4 kartiert werden.

Identifizierung der Gene für die Proteine TP1, TP2 und TP3 mit Hilfe der N-terminalen Aminosäuresequenz

Nachdem die viralen Strukturproteine TP1, TP2 und TP3 in sauberer und mengenmäßig ausreichender Form vorlagen, konnten sie mit Hilfe des Edman-Abbaus im Sequenator sequenziert werden. Die Sequenzierung auf Proteinebene wurde

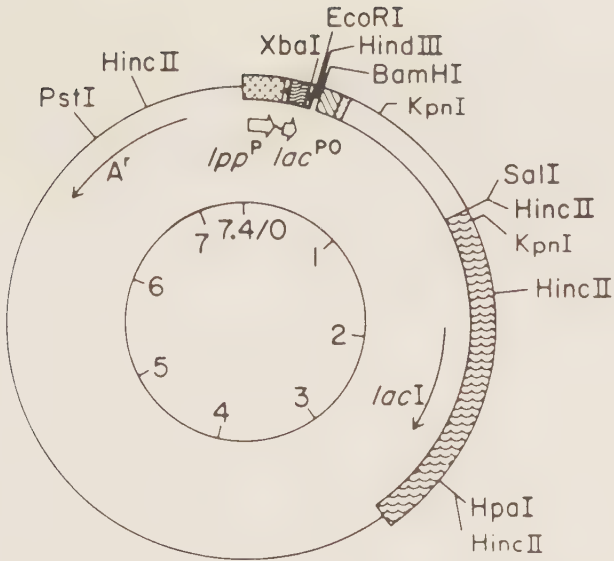


Abb. 14 : Der Expressionsvektor pINI11.

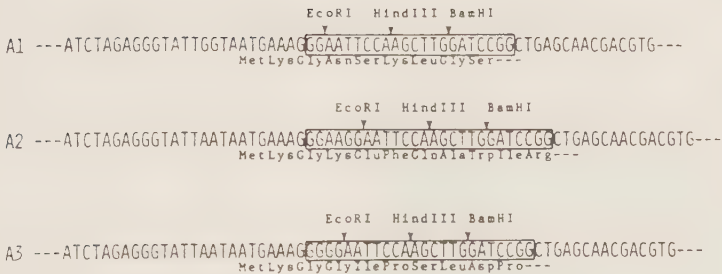


Abb. 15 : Die Klonierungsstellen in den drei Leserahmen der Vektoren pINI11 A1-3.

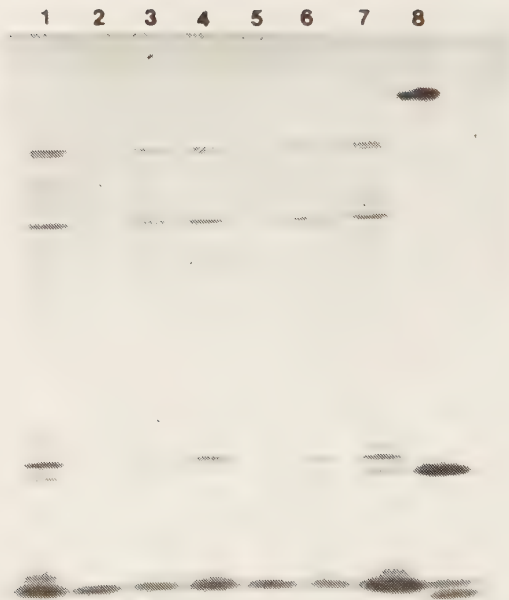


Abb. 16 : In *E. coli* exprimiertes Strukturprotein TP2 von TTV1. SDS-Polyacrylamid-Gel positiver Klone (1)-(7), (8) TTV1.



Abb. 17 : Western-Blot des Gels aus Abb. 16. Nachweis der Expression von TP2 in *E. coli* Klonen mit Hilfe von Antikörpern gegen das Protein.

```

      GAGAGAATCATGATAATAATCAAGACCAAAAATAGAGAATATACGATAGAT
1  -----+-----+-----+-----+-----+-----+-----+
      MetIleIleIleLysThrLysAsnArgGluTyrThrIleAsp

      GAGAATAAGGAGATAAATTCCATAAGTGCAGCTGATACTATATGTCCATAT
52  -----+-----+-----+-----+-----+-----+-----+
      GluAsnLysGluIleAsnSerIleSerAlaAlaAspThrIleCysProTyr

      CGTGCATATCGGTCGTATCATAGAAAAATCCCAATAAAAATCGACGATTGT
103 -----+-----+-----+-----+-----+-----+-----+
      ArgAlaTyrArgSerTyrHisArgLysIleProIleLysIleAspAspCys

      AAAGAGGCGAGAAAGATAGCTGGGAGCATGACGCATTACGTTCATATTGCCA
154 -----+-----+-----+-----+-----+-----+-----+
      LysGluAlaArgLysIleAlaGlySerMetThrHisTyrValIleLeuArg

      CAGCTAAAGGAGCAGGGATGTGAGACGGAAAAGATAGTTGACCCAGAGCTG
205 -----+-----+-----+-----+-----+-----+-----+
      GlnLeuLysGluGlnGlyCysGluThrGluLysIleValAspProGluLeu

      AAGTTTAGAGCTGATGCTATATGTGATGGGGACGCATTGTTGAGATTAAGA
256 -----+-----+-----+-----+-----+-----+-----+
      LysPheArgAlaAspAlaIleCysAspGlyAspAlaLeuLeuArgLeuArg

      CAAACTTCAATCGGGACTGGCATGCGCTTAATAAGTGGCAATTGAAAGTGT
307 -----+-----+-----+-----+-----+-----+-----+
      GlnThrSerIleGlyThrGlyMetArgLeuIleSerGlyAsnEnd

```

Abb. 18 : Die kodierende Region für das Protein TP1.
DNA-Sequenz des Gens und Aminosäuresequenz des Proteins.

von R. Hensel und F. Lottspeich durchgeführt. Da die Amino-
enden der Proteine nicht blockiert waren, gelang es, im
Sequenator eine Anzahl von Aminosäuren vom N-Terminus her zu
identifizieren. Der Vergleich der DNA-Sequenz mit den
N-terminalen Aminosäuresequenzen der drei TTV1-Struktur-
proteine ergab eine eindeutige Identifizierung der
korrespondierenden Gene. Nur je eine mögliche DNA-Sequenz
wurde gefunden, wenn alle sechs Leseraster des TTV1-Genoms
auf die Fähigkeit hin untersucht wurden, die N-terminalen

Aminosäuresequenzen der Proteine TP1, TP2 und TP3 zu kodieren. Die Positionen der Gene für TP1 und TP2 entsprachen dabei der Kartierung mit Hilfe des Expressionsvektors pIN III.

Die Organisation der Gene von TP1, TP2 und TP3

Das Gen für TP1 entspricht dem offenen Leserahmen Nr. 6 (s. Tab. 4), ist 339 Basenpaare lang und kodiert für 113 Aminosäuren (Abb. 18), das Gen für TP2 entspricht dem offenen Leserahmen Nr. 19, ist 417 bp lang und kodiert für 139 Aminosäuren (Abb. 19). Das Gen für TP3 entspricht dem offenen Leserahmen Nr. 21, ist 480 bp lang und kodiert für 160 Aminosäuren (Abb. 20). Die Gene für TP1, TP2 und TP3 liegen auf benachbarten Restriktionsfragmenten (Abb. 21) und kodieren alle drei auf dem gleichen Strang. Das Gen für TP1 ist 2.192 Basenpaare vom Gen für TP2 und dieses 564 bp vom Gen für TP3 entfernt, ein Umstand, der eine Co-Transkription aller drei Gene und damit eine Operonstruktur nahezu ausschließt.

```

1      GTTAATGCTATGAAAGTACTTAGGTTAGGATTCACCGATAAATATATCGAG
      -----+-----+-----+-----+-----+-----+
      MetLysValLeuArgLeuGlyPheThrAspLysTyrIleGlu

152    AAGCAGGAGCCTATCGATGTAGTATTCGATAAGTACCAGCCGCTGGATAT
      -----+-----+-----+-----+-----+-----+
      LysGlnGluProIleAspValValPheAspLysTyrGlnProLeuGlyTyr

103    GTAGCTGTAGTTGAATTACCTTCGAGGATACCGTGGGATAATCGAGATTCAG
      -----+-----+-----+-----+-----+-----+
      ValAlaValValGluLeuProSerArgIleProTrpIleIleGluIleGln

154    AGGAGGGAATGGATAGAAAGGTTCAACAATGCCGCGTGACATATTTAGA
      -----+-----+-----+-----+-----+-----+
      ArgArgGluTrpIleGluArgPheIleThrMetProArgAspIlePheArg

205    GAATTATCATTTGACATAATTATACTGAGGAGAAAGCTGGAGCCTACTCCT
      -----+-----+-----+-----+-----+-----+
      GluLeuSerPheAspIleIleIleLeuArgArgLysLeuGluProThrPro

256    CAATATCGTATGATAAGGGATATAGTGTGAGATTTAAGACAGTCAAGAGGA
      -----+-----+-----+-----+-----+-----+
      GlnTyrArgMetIleArgAspIleValSerAspLeuArgGlnSerArgGly

307    TATGCGAGCGGGATGGTGATACTGCCTAATGGACTAACATATGATGGGGAC
      -----+-----+-----+-----+-----+-----+
      TyrAlaSerGlyMetValIleLeuProAsnGlyLeuThrTyrAspGlyAsp

358    CTGCTTGAAGGTATAGAAGTTATGGAAGGAGTTGATGTAATAGCCTATACG
      -----+-----+-----+-----+-----+-----+
      LeuLeuGluGlyIleGluValMetGluGlyValAspValIleAlaTyrThr

409    TTGGGTCTAATCGATTCTAACAGATAAGTATATTAATACCTAATCATTTT
      -----+-----+-----+-----+-----+-----+
      LeuGlyLeuIleAspPheEnd

```

Abb. 19 : Die kodierende Region für das Protein TP2
DNA-Sequenz des Gens und Aminosäuresequenz des Proteins.

```

1  ATTAGGACAATGGTTGAGATTAACTAGTTAATAAGGAAATAATAAAGTTT
   -----+-----+-----+-----+-----+-----+
   MetValGluIleLysLeuValAsnLysGluIleIleLysPhe

52  GGATTGGCACTTGAATTGTGAATGAGCTTAATGAATACCTATACGGCGCA
   -----+-----+-----+-----+-----+-----+
   GlyLeuAlaLeuGlyIleValAsnGluLeuAsnGluTyrLeuTyrAlaAla

103  ATGCCTCCAATGTATGACATATTATCTAAATTGTACGGATATGGAAGGACT
   -----+-----+-----+-----+-----+-----+
   MetProProMetTyrAspIleLeuSerLysLeuTyrGlyTyrGlyArgThr

154  GTAAATGCAATGCTGTATAGTGTGCTGTTAATTTACTCGAATCGAGGATT
   -----+-----+-----+-----+-----+-----+
   ValAsnAlaMetLeuTyrSerValLeuPheAsnLeuLeuGluSerArgIle

205  GATACACTGAGAAGACTTGACCTTAGTAAGTTCTTTGCTGTGGTACATAT
   -----+-----+-----+-----+-----+-----+
   AspThrLeuArgArgLeuAspLeuSerLysPhePheAlaValValAlaTyr

256  ATGGATGCTGTTAAAGATACCGTTGAGCTTGCAATAAATGGATATGCATAT
   -----+-----+-----+-----+-----+-----+
   MetAspAlaValLysAspThrValGluLeuAlaIleAsnGlyTyrAlaTyr

307  GTTACGACTTCTGGATTTACAGTGTATCCTGCATCACAGATTACAGGAACG
   -----+-----+-----+-----+-----+-----+
   ValThrThrSerGlyPheThrValTyrProAlaSerGlnIleThrGlyThr

358  TATGAGAGCGACTTCAGTCAAGCTAAGGCAGTACCCGCAAATACTACAGCA
   -----+-----+-----+-----+-----+-----+
   TyrGluSerAspPheSerGlnAlaLysAlaValProAlaAsnThrThrAla

409  ACACAGGTGTGGTTGCACCCAAGAAGTTTGCTTTGATTACACAGAATAAG
   -----+-----+-----+-----+-----+-----+
   ThrGlnValTrpPheAlaProLysLysPheAlaLeuIleThrGlnAsnLys

460  ATATATACATTCCCTAGCTCCTTATTGGTTCTAAACTAACTAAAGCTTATT
   -----+-----+-----+-----+-----+-----+
   IleTyrThrPheLeuAlaProTyrTrpPheEnd

```

Abb. 20 : Die kodierende Region für das Protein TP3.
DNA-Sequenz des Gens und Aminosäuresequenz des Proteins.

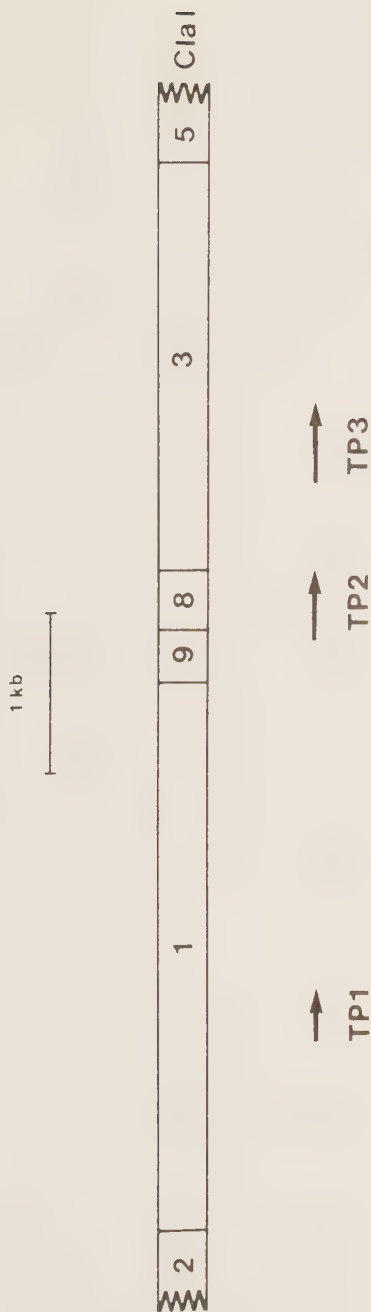


Abb. 21 : Lokalisierung der Gene für TP1, TP2 und TP3 auf dem TTV1-Genom (ClaI-Restriktionsmuster).

Kodon-Benutzung der Proteine TP1, TP2 und TP3

Bis auf die Kodons AAC für Asparagin, TGC für Cystein, CAC für Histidin und dem Stopkodon TAG werden alle übrigen Kodons in den drei Strukturproteinen verwendet (Tab. 5). Im Durchschnitt kommen Kodons mit den Basen A oder T als dritte Base des Kodons öfter vor und Kodons mit C als dritter Base werden am wenigsten häufig verwendet. Dieses Ergebnis war durch den niedrigen GC-Gehalt der TTV1-DNA zu erwarten.

Ribosomenbindungsstellen

Ein Vergleich des 3'-Endes der *Thermoproteus tenax* 16S rRNA mit den Sequenzen oberhalb der für die Proteine TP1, TP2 und TP3 kodierenden Gene zeigte, daß keine Shine-Dalgarno Sequenzen vor den Startkodons der Gene lagen. Daraufhin wurden die "upstream" Sequenzen aller offenen Leserahmen untersucht, aber auch dort konnte keine dem 3'-Ende der 16S rRNA komplementäre Sequenz als Ribosomenbindungsstelle gefunden werden.

Amino- säure	Codon	TP1	TP2	TP3	Gesamt
Gly	GGG	3	2	—	5
Gly	GGA	1	5	7	13
Gly	GGT	—	2	—	2
Gly	GGC	2	—	—	2
Glu	GAG	6	4	4	14
Glu	GAA	2	7	3	12
Asp	GAT	5	8	3	16
Asp	GAC	3	3	3	9
Val	GTG	—	2	5	7
Val	GTA	—	6	3	9
Val	GTT	1	3	5	9
Val	GTC	1	—	—	1
Ala	GCG	1	1	1	3
Ala	GCA	3	—	11	14
Ala	GCT	4	1	5	10
Ala	GCC	—	1	—	1
Arg	AGG	—	7	2	9
Arg	AGA	6	4	2	12
Ser	AGT	2	—	3	5
Ser	AGC	1	1	1	3
Lys	AAG	6	3	7	16
Lys	AAA	4	2	3	9
Asn	AAT	4	1	8	13
Asn	AAC	—	—	—	—
Met	ATG	3	5	5	13
Ile	ATA	12	12	5	29
Ile	ATT	—	2	5	7
Ile	ATC	4	4	—	8
Thr	ACG	3	1	2	6
Thr	ACA	—	2	7	9
Thr	ACT	3	1	3	7
Thr	ACC	1	1	1	3

Tab. 5 : Kodonbenützung der drei Strukturproteine
TP1, TP2 und TP3.

Trp	TGG	-	2	2	4
Opal	TGA	1	-	-	1
Cys	TGT	4	-	-	4
Cys	TGC	-	-	-	-
Amber	TAG	-	-	-	-
Ochre	TAA	-	1	1	2
Tyr	TAT	4	6	10	20
Tyr	TAC	1	1	3	5
Leu	TTG	3	1	3	7
Leu	TTA	2	4	2	8
Phe	TTT	1	2	6	9
Phe	TTC	-	4	4	8
Ser	TCG	1	1	1	3
Ser	TCA	1	3	1	5
Ser	TCT	-	-	2	2
Ser	TCC	1	-	-	1
Arg	CGG	1	-	-	1
Arg	CGA	1	-	-	1
Arg	CGT	1	2	-	3
Arg	CGC	1	-	-	1
Gln	CAG	2	4	3	9
Gln	CAA	1	1	1	3
His	CAT	2	-	-	2
His	CAC	-	-	-	-
Leu	CTG	1	4	3	8
Leu	CTA	1	2	3	6
Leu	CTT	-	3	5	8
Leu	CTC	-	-	1	1
Pro	CCG	-	3	-	3
Pro	CCA	3	-	1	4
Pro	CCT	-	5	3	8
Pro	CCC	-	-	2	2

Mögliche Promoterelemente

Der Vergleich der Sequenz in der Umgebung der Startkodons der Gene für die Strukturproteine TP1, TP2 und TP3 mit vermutlichen Promoterstrukturen in anderen Archaeobakterien, zeigte eine große Ähnlichkeit mit der von Wich *et al.* (1987) beschriebenen Promotersequenz für die Expression stabiler RNA-Gene in *Thermoproteus tenax*. Allerdings sind die Distanzen der AT-reichen Sequenzen zu den Startkodons der Gene sehr unterschiedlich. Im Falle von TP1 beträgt der

- (1) CAAAGAGGTTTAAAGATAAGATATTGGACAAGCGTAATCTAACAATAAAA

 TAGAAAAATCAGAGAATACAGGAGAGAATCATGATAATAATCAAGACCAA
 MetIleIleIleLysThrLys

- (2) TGAAGTAGACATAATGGCCCAGTATATGGCCATTTATATTATGAGCAAAAG

 AGATAAAGAGTTGATGAGTGCCTTAATGCTATGAAAGTACTTAGGTTAGGA
 MetLysValLeuArgLeuGly

- (3) AGGTTAAATTTAAGGAAAAGATATACTTTTATACCGCCATTTAGGGAGAAAAG

 ATTTAAATACCATGGATAGTAATTAGGACAATGGTTGAGATTAACTAGTT
 ***** MetValGluIleLysLeuVal

Abb. 22 : "upstream" Sequenzen der Gene für die Proteine (1) TP1, (2) TP2 und (3) TP3. Mögliche Promoterstrukturen sind mit * gekennzeichnet.

Abstand 66 bp und bei TP2 42 bp. Nur im Falle von TP3 wurde in einer erwarteten Distanz (23 bp) vor dem Startkodon eine AT-reiche Sequenz (TTTAAA) gefunden, die mit dem für *Thermoproteus tenax* beschriebenen "box A"-Promoterelement identisch ist. Die für den Transkriptionsstart für stabile RNA-Gene charakteristische "box B", xGC, konnte in keinem Fall im richtigen Abstand zur "box A" gefunden werden (Abb. 22).

Eigenschaften der Strukturproteine TP1, TP2 und TP3

Nach der DNA-Sequenzanalyse und der zugehörigen Aminosäuresequenz beträgt das Molekulargewicht von TP1 14,8 kd, von TP2 18,8 kd und von TP3 18,9 kd. Die Größen der Proteine wurden von der DNA-Sequenz abgeleitet unter der Voraussetzung, daß sich keine Introns in den Genen befinden und die Proteine an ihren Carboxylenden nicht prozessiert werden. Die Unterschiede zu den im SDS-Polyacrylamid-Gel ermittelten Molekulargewichten ergeben sich wahrscheinlich aus der Konfiguration der Proteine und damit einem gegenüber ihrem apparenten Molekulargewicht geänderten Laufverhalten.

TP1 reagiert auf Grund von 19% basischer Aminosäuren relativ basisch, TP2 reagiert leicht sauer und TP3 verhält sich neutral (Tab. 6). Die geladenen Aminosäuren treten bei TP2 im Zentrum der Sequenz gehäuft auf, bei den anderen Proteinen sind sie gleichmäßig über die Sequenz verteilt.

Aminosäure	Vorkommen in		
	TP1	TP2	TP3
Ala	8	3	17
Arg	10	13	4
Asn	4	1	8
Asp	8	11	6
Cys	4	-	-
Gln	3	5	4
Glu	8	11	7
Gly	6	9	7
His	2	-	-
Ile	16	18	10
Leu	7	14	17
Lys	10	5	10
Met	3	5	5
Phe	1	6	10
Pro	3	8	6
Ser	6	5	8
Thr	7	5	13
Trp	-	2	2
Tyr	5	7	13
Val	2	11	13

Tab. 6 : Aminosäurezusammensetzung der Strukturproteine TP1, TP2 und TP3.

Die Proteine TP1 und TP2 sind an die DNA gebunden (Janekovic *et al.*, 1983). Alle drei Proteine reagieren hydrophob. TP3 besteht zu 50% aus hydrophoben Aminosäuren und in seiner zentralen Region findet sich eine extrem hydrophobe Sequenz von 26 Aminosäureresten.

Die Proteine besitzen untereinander weder auf der DNA- noch auf der Aminosäureebene eine Ähnlichkeit, ein Ergebnis, das auch durch negative Kreuzhybridisierungsexperimente mit Antikörpern gegen die drei Strukturproteine bestätigt wurde.

DIE GENOMISCHEN VARIANTEN VON TTV1

Die erste spontan auftretende genomische Variante von TTV1 wurde in dem *Thermoproteus tenax* Klon 10/14 entdeckt. Dieser Klon trägt das gegen das ClaI-Fragment Nr. 5 hybridisierende DNA Stück. Später traten sowohl in infizierten als auch in "lysogenen" Kulturen, die dieses hybridisierende DNA-Stück nicht besaßen, weitere Varianten auf, so daß das hybridisierende DNA-Stück mit den Varianten in keinem direkten Zusammenhang stehen muß. Bis jetzt wurden fünf solcher genomischer Varianten entdeckt. Im phänotypischen Erscheinungsbild konnte mit Hilfe elektronenmikroskopischer Analyse kein Unterschied zwischen diesen Varianten und dem Wildtyp-Virus festgestellt werden. Auch in der Virus-Wirt-Wechselbeziehung ergaben sich im Wachstum, in der Reinfizierung und im Lyseverhalten keine Unterschiede. Die Virusvarianten traten entweder rein oder als Mischung zweier Varianten im Lysat auf (Abb. 23). Das Mischungsverhältnis variierte dabei zwischen 10% der einen und bis zu 90% der anderen Variante.

Struktur der Variantengenome

Die Restriktionsanalyse der Variantengenome zeigte, daß in jedem Fall der Variation die ClaI-Fragmente Nr. 4 und Nr. 6 involviert sind (Abb. 24). Die Varianten VT1, VT2 und VT5 lassen sich durch Deletionen in diesen ClaI-Fragmenten erklären. Im Falle von VT4 bewirkt eine Insertion oder eine

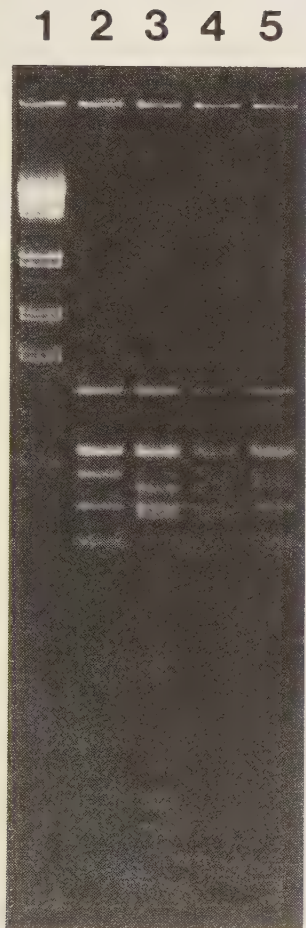


Abb. 23 : Genomische Varianten von TTV1. (1) HindIII Spaltung von T₅-DNA, ClaI-Spaltungsmuster von (2) TTV1-Wildtyp, (3) VT3, (4) Mischung von VT3 und VT4, (5) Mischung von VT5 und Wildtyp.

Verdoppelung eines bestimmten DNA-Abschnitts die Variation. In der Variante VT3 scheint dagegen die genomische Verschiedenheit vom Wildtyp-Virus von komplexerer Natur zu sein. Aus diesem Grund wurden die ClaI-Fragmente Nr. 4 und Nr. 6 der Variante VT3 sequenziert und einer genauen Analyse unterzogen.

Sequenzanalyse der ClaI-Fragmente Nr. 4 und Nr. 6 der genomischen Variante VT3

Die Sequenzierung der ClaI-Fragmente Nr. 4 und Nr. 6 der Variante VT3 und der Vergleich mit den ClaI-Fragmenten Nr. 4 und Nr. 6 des Wildtyps ergab, daß ganz bestimmte DNA-Sequenzen an der Bildung der Variation im Genom beteiligt sind. Diese DNA-Sequenzen haben untereinander eine außerordentlich große Ähnlichkeit in ihrer Sequenz (Abb. 25-28). Auffallend ist außerdem der hohe GC-Gehalt von 58% in diesen kurzen Sequenzen gegenüber einem GC-Gehalt von 37% im gesamten Genom. An den linken Flanken dieser DNA-Stücke findet sich jeweils die gleiche kurze DNA-Sequenz von 8 bp Länge mit dem Konsensus ACXCCTAC.

Bildung der Variante VT3

Um ein DNA-Fragment von 69 bp Länge, das auch im ClaI-Fragment Nr. 6 des Wildtyps vorkommt, gruppieren sich im ClaI-Fragment Nr. 6 der Variante VT3 fünf weitere

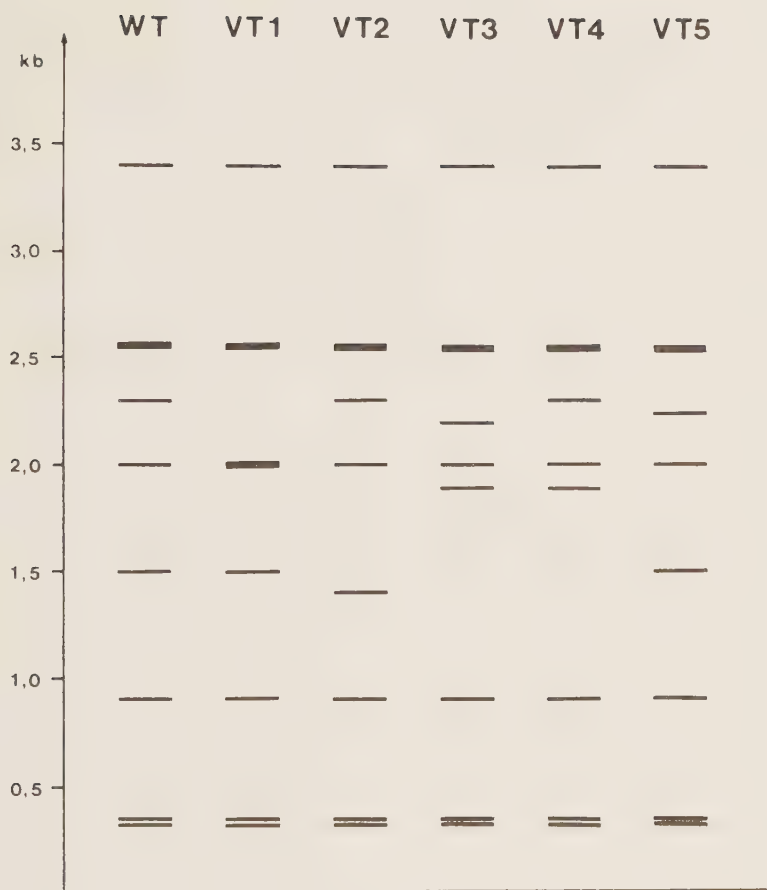


Abb. 24 : Genomische Varianten von TTV1.
 Die Fragmente entsprechen dem ClaI-
 Restriktionsmuster, (WT) entspricht dem
 Wildtyp-Virus.

DNA-Sequenzen die entweder aus dem ClaI-Fragment Nr. 4 des Wildtyps stammen (Abb. 29), oder in VT3 durch Duplikation dieser Sequenzen aus VT3-ClaI-4 entstanden sind. Es sind dies fünf kurze DNA-Stücke von zweimal 66 bp, 72 bp, 78 bp und 84 bp Länge. Dabei wurde im Vergleich zum ClaI-Fragment Nr. 4 des Wildtyps die DNA-Sequenz von 66 bp Länge verdoppelt, die sich nun in VT3 beiderseits des DNA-Stücks wiederfindet, das aus dem ClaI-Fragment Nr. 6 des Wildtyps stammt. VT3-ClaI-6 unterscheidet sich vom Wildtyp-Fragment zusätzlich noch durch eine Deletion einer kurzen Sequenz von 30 bp.

Das ClaI-Fragment Nr. 4 in der Variante VT3 unterscheidet sich vom ClaI-Fragment Nr. 4 im Wildtyp dagegen nur durch eine einfache Deletion einer DNA-Sequenz von 102 bp.

Auffallend an diesen DNA-Rearrangements ist, daß die Länge der involvierten DNA-Sequenzen jeweils durch drei teilbar ist, das bedeutet, daß unabhängig von der Anordnung dieser Sequenzen das Leseraster erhalten bleibt. Die kurze Konsensussequenz ACXCCTAC scheint dabei eine Zielsequenz zu sein, da sie in jeder Umwandlung involviert ist, aber nicht verdoppelt wird und an den linken Flanken der Elemente erhalten bleibt.

```

451  AAGTAGGAATTAATTAGCGTATCTGCAACTCCAACGCCTACGCCTACACC
      *****eeeeeeeeee
501  CACACCAACGCCAACTCCGACGCCTACTCCTACCCCTACTCCGACACCTA
      eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee
551  CGCCGACTCCCCTCCAACACCTACTCCGACTCCGACACCTACTCCAACA
      eeeeeeeeeeeeeeeee*****cccccccccccccccccccccc
601  CCAACGCCAACTCCGACACCGACTCCAACGCCAACGCCTACTCCTACACC
      ccccccccccccccccccccccccccccccc*****ssssssss
651  AACCCCAACACCTACCCCTACTCCTACACCTACGCCGACTCCGACTCCGA
      sssssssssssssssssssssssssssssssssssssssssssssss
701  CACCCACTCCTACGCCTACTCCGACTCCTACCCCAACTCCTACTCCCACT
      sssssssssss*****nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
751  CCCACCCCTACTCCGACTCCAACACCGACTCCGACACCTACTCCGACGCC
      nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn*****
801  TACTCCGACTCCAACGCCGACTCCTACACCAACGCCTTCACCTACACCAA
      ***aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
851  CTCCCACCCCTACTCCGACTCCGACGCCAACTCCGACTCCCCTCCAACG
      aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa***
901  CCTACACCAACTCCGACGTCATGTAAACCGATAATTACTACTGGAGTATC
      *****

```

Abb. 25 : Die in die Variation involvierte Sequenz im ClaI-Fragment Nr. 4 der Wildtyp-DNA.

```

451  AAGTAGGAATTAATTAGCGTATCTGCAACTCCAACGCCTACGCCTACACC
      *****eeeeeeeeee

501  CACACCAACGCCAACTCCGACGCCTACTCCTACCCCTACTCCGACACCTA
      eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee

551  CGCCGACTCCCACTCCAACACCTACTCCGACTCCGACACCTACTCCAACA
      eeeeeeeeeeeeeeeee*****cccccccccccccccccccccccc

601  CCAACGCCAACTCCGACACCGACTCCAACGCCAACGCCTACTCCTACACC
      ccccccccccccccccccccccccccccccccccccc*****ssssssss

651  AACCCCAACACCTACCCCTACTCCTACACCTACGCCGACTCCGACTCCGA
      sssssssssssssssssssssssssssssssssssssssssssssssss

701  CACCCACTCCTACGCCTACTCCGACTCCTACCCCAACTCCTACTCCCACT
      sssssssssss*****nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn

751  CCCACCCCTACTCCGACTCCAACACCGACTCCGACACCTACTCCGACGCC
      nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn*****      ****

801  TACACCAACTCCGACGTCATGTAAACCGATAATTACTACTGGAGTATCAT
      ***

```

Abb. 26 : Die in die Variation involvierte Sequenz im ClaI-Fragment Nr. 4 der Variante VT3.

Die 8 bp Konsensussequenz kommt im ganzen Genom von TTV1 insgesamt 35 mal vor, allerdings nur in den ClaI-Fragmenten Nr. 4 und Nr. 6. Im ClaI-Fragment Nr. 6 gibt es zwischen dem Nukleotid 10.077 und dem Nukleotid 10.695 insgesamt 11 dieser kurzen Sequenzen und im ClaI-Fragment Nr. 4 gibt es 24 dieser Sequenzen zwischen den Nukleotiden 11.527 und 11.941. Dies ergibt sich auch aus der außerordentlich großen Ähnlichkeit der Sequenzen innerhalb der ClaI-Fragmente Nr. 4 und Nr. 6.

```

951      TCTTCACCATCTCCTACTCCTACACCGACACCGACGCCTACGCCACACC
                                     *****VVVVVVVVVV
1001     GACATATGATATCACTTATGTAGTATTTGATGTAACGCCTTCACCTACTC
VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV
1051     CAACTCCTACTCTTACATCGACACCCACTCCTACTCCTACGCCAACATAT
VV*****OOOOOOOOOOOOOOOOOOOOOOOOOO*****

```

Abb. 27 : Die in die Variation involvierte Sequenz im ClaI-Fragment Nr. 6 der Wildtyp-DNA.

951 TCTTACCATCTCTCTACTCTCTACACCGACACCAACGCCTACGCTTACACC
*****eeeeeeeeee

1001 CACACCAACGCCAACTCCGACGCCTACTCTTACCCCTACTCCGACACCTA
ee

1051 CGCCGACTCCCCTCCAACACCTACTCCGACTCCGACACCTACTCCAACA
eeeeeeeeeeeeeeeeee*****cccccccccccccccccccccccccccc

1101 CCAACGCCAACTCCGACACCGGACTCCAACGCCAACGCCTACTCCGACTC
cccccccccccccccccccccccccccccccccccc*****nnnnnnnn

1151 CTACCCCAACTCTACTCCCCTCCCACCCCTACTCCGACTCCAACACCG
nn

1201 ACTCCGACACCAACGCCTACGCCTACACCGACATATGATATCACTTATGT
nnnnnnnnnnnnnn*****vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv

1251 AGTATTTGATGTAACGCCTTCACCTACTCCAACACCTACTCCGACTCCGA
vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv*****cccccccccc

1301 CACCTACTCCAACACCAACGCCAACTCCGACACCGACTCCAACGCCAACG
cc***

1351 CCTACTCTTACACCAACCCCAACACCTACCCCTACTCTTACACCTACGCC
*****ss

1401 GACTCCGACTCCGACACCCACTCCTACTCCTACGCCAACATATGATATAA
ssssssssssssssssssssssssssss*****

Abb. 28 : Die in die Variation involvierte Sequenz im ClaI-Fragment Nr. 6 der Variante VT3.

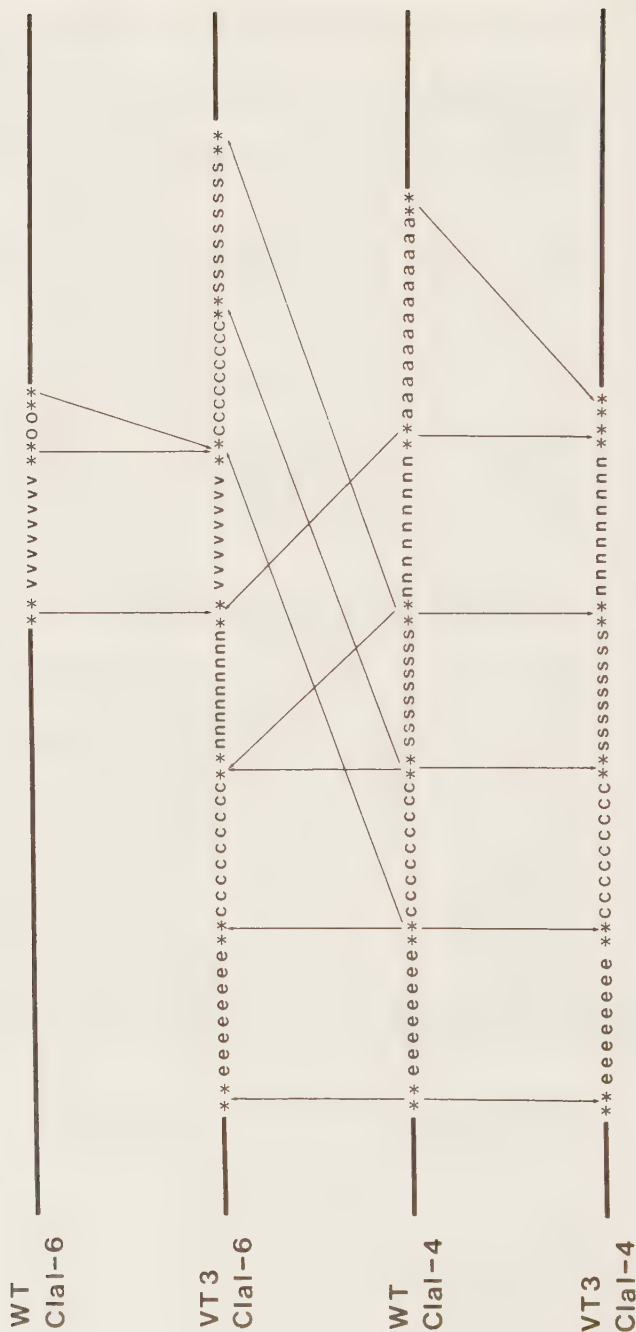


Abb. 29 : Bildung der Variante VT3 aus definierten DNA-Sequenzen der Wildtyp-DNA (Clal-Restriktionsmuster). Homologe Sequenzen sind gleich gekennzeichnet (siehe auch Abb. 25 - 28). (*) entspricht der Konsensussequenz.

DAS PROTEIN TPX

Die Sequenzierung des ClaI-Fragments Nr. 6, eines der beiden Stücke des TTV1-Genoms, das in der Variantenbildung involviert ist, ergab einen seltsamen offenen Leserahmen mit einer dreifachen Wiederholung einer 22 Aminosäuren langen Sequenz (Abb. 30). Die Sequenz beginnt mit Val-Thr gefolgt von 10 Paaren Pro-Thr mit einigen, funktionell äquivalenten Aminosäureaustauschen. Diese Sequenzen sind voneinander getrennt durch zwei Kopien einer spezifischen Sequenz von neun Aminosäuren. Die C-terminale Kopie unterscheidet sich von der oberhalb gelegenen nur durch den Austausch eines Valins gegen ein Isoleucin. Auf der DNA-Ebene zeigen diese 27 bp Wiederholungen nur drei Austausche in den Nukleotiden, von denen einer, der zur Substitution von Valin durch Isoleucin führt, nicht in der dritten, sondern in der ersten Position des Kodons ist. Die starke Konservierung der Basen selbst in der dritten Position des Kodons legt eine bestimmte, bis jetzt unbekannte Funktion auf der DNA-Ebene nahe.

```

1  CCATTAGTAATGTACTTATCAATAAATGGAAGTACTGCAAACGTAAAGGTA
   MetTyrLeuSerIleAsnGlySerThrAlaAsnValLysVal

52  TACAAACAGGGCTCAAACATTGGAAGTGTAAAGTGGTAACTATTCTACTACT
   TyrLysGlnGlySerAsnIleGlyThrValSerGlyAsnTyrSerThrThr

103 CCATATGGAAATCCATCTATGGCAGGATATGGAACAGTTGATAAGCATTAT
   ProTyrGlyAsnProSerMetAlaGlyThrGlyThrValAspLysHisTyr

154 GCTAACTTTATTGTGTTACCATATGAGCCAGACCCACAGGTAACAGTTACT
   AlaAsnPheIleValLeuProTyrGluProAspProGlnValThrValThr
   -----

205 CCTATATCTTCACCATCTCCTACTCCTACACCGACACCGACGCTACGCCT
   ProIleSerSerProSerProThrProThrProThrProThrProThrPro
   -----

256 ACACCGACATATGATATCACTTATGTAGTATTTGATGTAACGCCTTCACCT
   ThrProThrTyrAspIleThrTyrValValPheAspValThrProSerPro
   -----*****-----

307 ACTCCAACCTCCTACTCTTACATCGACACCCACTCCTACTCCTACGCCAACA
   ThrProThrProThrLeuThrSerThrProThrProThrProThrProThr
   -----

358 TATGATATAACTTATGTAATATTCGATGTAACACCTTCACCCACACCTACT
   TyrAspIleThrTyrValIlePheAspValThrProSerProThrProThr
   *****-----

409 CCAACACCTACTCCTACACCAACTCCGACGCCAACTCCTACATCGACAACG
   ProThrProThrProThrProThrProThrProThrProThrSerThrThr
   -----

460 TCATCTAACATCTAGTTGTAATCCAATAATAACAACAGGTATTCAATTCCT
   SerSerAsnIleEnd

```

Abb. 30 : Die kodierende Region für das Protein TPX.
DNA-Sequenz des Gens und Aminosäuresequenz des
Proteins. Markierte Sequenzen sind annähernd homolog.

in vitro Synthese eines Peptids von TPX

Um die mögliche Expression des Proteins TPX nachzuweisen, wurde ein Teil dieses Proteins *in vitro* synthetisiert. Dafür bot sich im Zentrum des Proteins ein Teil der hochrepetitiven Pro-Thr Sequenz und eine der Wiederholungen der neun Aminosäuren an. An das iso-1-Cytochrom C wurde das Peptid β -Ala-(Thr-Pro)₃-Thr-Tyr-Asp-Ile-Thr-Tyr-Val-Val-Phe-Asp-Val-Thr-Pro-Ser-Pro-Thr-ProOH gekoppelt. Die Peptidsynthese wurde nach Standardverfahren von E. Wünsch durchgeführt. Mit Hilfe dieses *in vitro*-Peptids wurden Kaninchen immunisiert und Antikörper gegen dieses Peptid gewonnen.

Identifizierung von TPX in der Wirtszelle

Mit Hilfe der "Western Blot" Analyse wurden im SDS-Polyacrylamid-Gel aufgetrennte Proteine von TTV1 gegen die Antikörper des *in vitro*-Peptids getestet. Das verwendete ELISA-Testverfahren ergab dabei keine Reaktion der Virusproteine gegen die Antikörper. Daraus läßt sich schließen, daß das Protein TPX nicht im Virus vorkommt und damit kein integraler Bestandteil des Virus ist.

Daraufhin wurde die Gesamtprotein-Fraktion von *Thermoproteus tenax* Klonen im SDS-Polyacrylamid-Gel aufgetrennt und ebenfalls einem ELISA-Test gegen die Antikörper des *in vitro*-Peptids unterzogen. Es zeigte sich,



Abb. 31 : Nachweis der Expression von TPX in *Thermoproteus tenax* Zellen. Western Blot Analyse verschiedener Zellklone.
 (1), (2) und (4) virusfreie Klone,
 (3), (5)-(8) virustragende Klone,
 (9) *E. coli*, (10) Antigen.

daß alle Klone, die einmal Virus produziert hatten, gegen die Antikörper reagierten (Abb. 31). In einem Fall zeigten sich zwei reagierende Banden, wobei die zweite Bande auch ein Abbauprodukt von TPX sein kann. Die gegen die Antikörper reagierenden Proteine zeigten im Gel ein Laufverhalten, das im Molekulargewicht einem Dimer von TPX entsprechen würde. Kurierte und virusfreie Zellklone reagierten nicht gegen die Antikörper. Das Protein TPX wird also nur in virustragenden oder virusinfizierten Zellen produziert. Die Funktion des Proteins TPX ist weiterhin unbekannt. Da das Protein kein integraler Bestandteil des Virus ist, aber in der Wirtszelle nachgewiesen werden kann, liegt der Schluß nahe, daß das Protein TPX in irgendeiner Weise in die Virus-Wirt-Wechselbeziehung involviert ist und vielleicht eine Art Regulationsprotein darstellt.

Das Protein TPX in der Variante VT3

Auf Grund der gewonnenen Erkenntnisse wurde auch die DNA-Sequenz des ClaI-Fragments Nr. 6 der Variante VT3 auf den offenen Leserahmen für das Protein TPX untersucht. Der offene Leserahmen umfaßt hier 486 bp gegenüber 462 bp im ClaI-Fragment Nr. 6 vom Wildtyp, was einem Protein von 162 Aminosäuren Länge entspricht. Die beiden Proteine sind in den ersten 85 Aminosäureresten identisch und variieren nur im Rest der Sequenz. Im Fall der Variante VT3 zeigte sich, daß das Protein TPX eine sehr lange Pro-Thr Sequenz

aufweist, in der 26 Pro-Thr Paare hintereinander liegen (Abb. 32).

```

1  CCATTAGTAATGTACTTATCAATAAATGGAAGTACTGCAAACGTAAAGGTA
   MetTyrLeuSerIleAsnGlySerThrAlaAsnValLysVal
   -----
52  TACAAACAGGGCTCAAACATTGGAAGTGTAAAGTGGTAACTATTCTACTACT
   TyrLysGlnGlySerAsnIleGlyThrValSerGlyAsnTyrSerThrThr
   -----
103 CCATATGGAAATCCATCTATGGCAGGATATGGAACAGTTGATAAGCATTAT
   ProTyrGlyAsnProSerMetAlaGlyTyrGlyThrValAspLysHisTyr
   -----
154 GCTAACTTTATTGTGTTACCATATGAGCCAGACCCACAGGTAACAGTTACT
   AlaAsnPheIleValLeuProTyrGluProAspProGlnValThrValThr
   -----
205 CCTATATCTTCACCATCTCCTACTCCTACACCGACACCAACGCCTACGCCT
   ProIleSerSerProSerProThrProThrProThrProThrProThrPro
   -----
256 ACACCCACACCAACGCCAACTCCGACGCCTACTCCTACCCCTACTCCGACA
   ThrProThrProThrProThrProThrProThrProThrProThrProThr
   -----
307 CCTACGCCGACTCCCACTCCAACACCTACTCCGACTCCGACACCTACTCCA
   ProThrProThrProThrProThrProThrProThrProThrProThrPro
   -----
358 ACACCAACGCCAACTCCGACACCGGACTCCAACGCCAACGCCTACTCCGAC
   ThrProThrProThrProThrProAspSerAsnAlaAsnAlaTyrSerAsp
   -----
409 TCCTACCCCAACTCCTACTCCCACTCCCACCCCTACTCCGACTCCAACACC
   SerTyrProAsnSerTyrSerHisSerHisProTyrSerAspSerAsnThr
   -----
460 GACTCCGACACCAACGCCTACGCCTACACCGACATATGATATCATTATGT
   AspSerAspThrAsnAlaTyrAlaThrAspIleEnd

```

Abb. 32 : Die kodierende Region für das Protein TPX im ClaI-Fragment Nr. 6 der genomischen Variante VT3. Die homologe Sequenz zum Protein TPX im Wildtyp ist unterstrichen.

Sekundärstruktur von TPX

Die Sekundärstruktur von TPX wurde nach der Methode von Chou und Fasman (1978) mit Hilfe des Computers errechnet. Die Proteine TPX im Wildtyp und in der Variante VT3 lassen sich in je drei Domänen unterteilen. Die NH_2 -terminalen Aminosäuren von 1 bis 85 sind in beiden Proteinen gleich. Danach folgen in WT-TPX drei kurze hyperflexible Bereiche von Pro-Thr Sequenzen die durch je neun Aminosäuren von gleicher Sequenz getrennt sind. Die dritte Domäne bilden schließlich die C-terminalen Aminosäuren. In VT3-TPX gibt es dagegen in der zweiten Domäne nur eine einzige lange Sequenz von 26 Pro-Thr Paaren hintereinander und eine weitere lange Sequenz von 38 verschiedenen Aminosäuren am C-Terminus.

Dieser möglichen Sekundärstruktur nach scheint es eher unwahrscheinlich, daß TPX eine enzymatische Funktion hat, vor allem, da sich die langen helikalen Pro-Thr Sequenzen in den Varianten ständig ändern.

Computervarianten von TPX

Da ein Teil des Proteins TPX von DNA-Sequenzen kodiert wird, die für die Variantenbildung verantwortlich sind, wurde untersucht, ob eine andere Anordnung dieser DNA-Sequenzen auf Proteinebene ähnliche Proteine wie TPX ergeben. Mit Hilfe des Computers wurden die in die

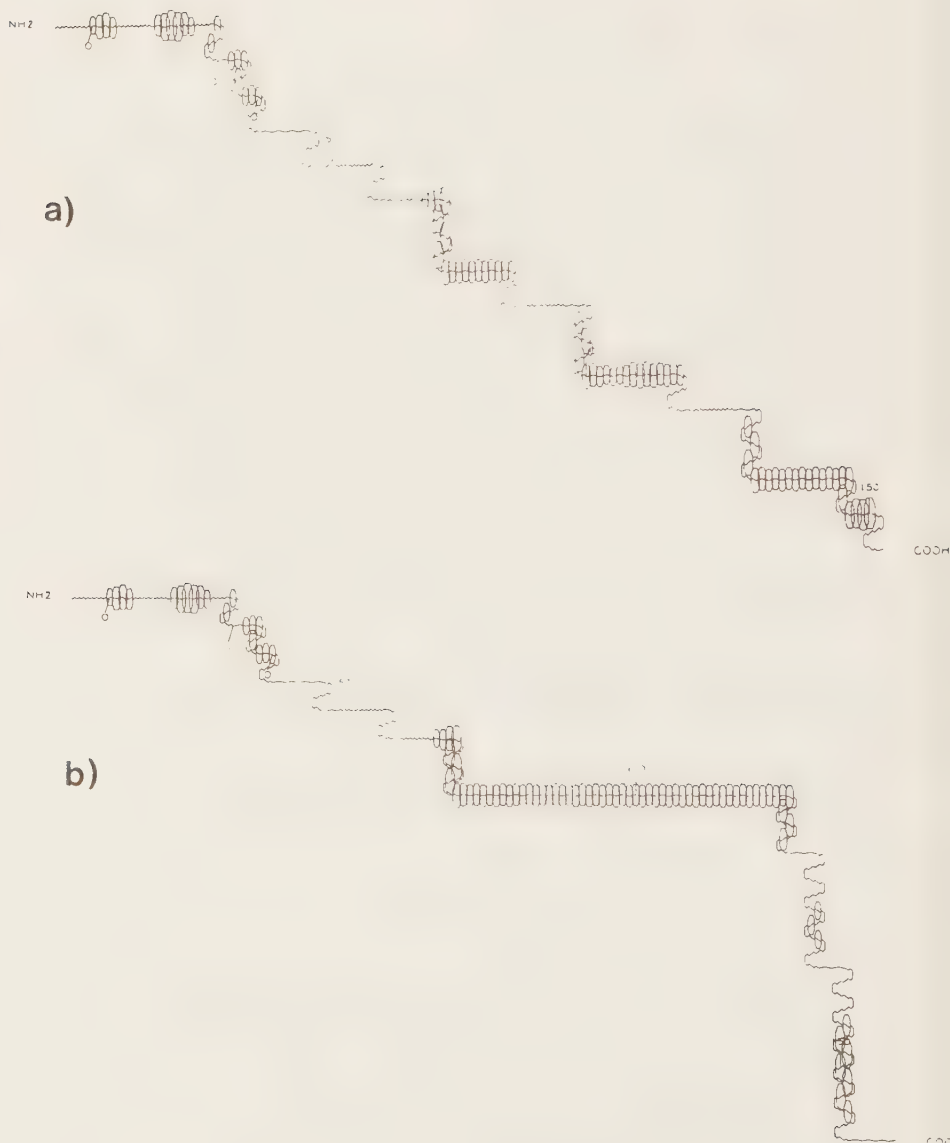


Abb. 33 : Schematisches Diagramm der möglichen
 Sekundärstruktur von TPX. a) TPX im Wildtyp,
 b) TPX in der Variante VT3.
 (0) hyperflexible Bereiche des Proteins.

Variantenbildung involvierten Sequenzen in unterschiedlichster Weise angeordnet, wobei auch mögliche Duplikationen und Deletionen berücksichtigt wurden. Die DNA-Sequenz, die die ersten 78 Aminosäuren von TPX kodiert und nicht in die Variantenbildung verwickelt ist, blieb dabei konserviert. Alle diese "Computer-Proteine" zeigten nun, ohne Rücksicht auf die Anordnung der DNA-Sequenzen, als Gemeinsamkeit längere Pro-Thr Sequenzen, die im Extremfall bis zu 62 Pro-Thr Paare hintereinander betrugen. Das läßt darauf schließen, daß diese Pro-Thr Sequenzen für die Funktion des Proteins TPX essentiell sind.

V. DISKUSSION

Das Virus TTV1

TTV1 war das erste Virus, das in den archaebakteriellen *Thermoproteales* entdeckt wurde. Es gehört zur Familie der stäbchenförmigen *Thermoproteus tenax* Viren, die sich in ihrer Morphologie stark ähneln und alle den gleichen Wirt infizieren können, obwohl sie untereinander sowohl auf DNA-, als auch auf Proteinebene keinerlei Verwandtschaft besitzen.

Das Genom des Virus, eine doppelsträngige lineare DNA von 15,9 kb, eignet sich als definierte genetische Einheit sehr gut zur Analyse von Genom- und Genorganisation. Solche überschaubaren Genome sind in *Thermoproteus tenax* auch noch bei den Viren TTV2, TTV3 und TTV4 vorhanden, die von Zillig *et al.* (1983, 1986) aus dem gleichen *Thermoproteus*-Stamm isoliert wurden, aber noch nicht näher charakterisiert sind.

Die TTV1-DNA hat überraschenderweise einen sehr niedrigen GC-Gehalt von nur 37%. Der Schmelzpunkt der DNA liegt demnach bei 83,8°C, während das Temperaturoptimum des Wirtsbakteriums *Thermoproteus tenax* bei 88°C liegt. Auch bei einem anderen Vertreter des thermophilen Zweiges der Archaebakterien, dem virusähnlichen Partikel SSV1 von *Sulfolobus spec. B12*, findet man mit 40% einen relativ geringen GC-Gehalt (Reiter *et al.*, 1987), während z.B. der GC-Gehalt des *Halobacterium*-Phagen Φ H 65% beträgt, obwohl

dieses Bakterium bei der vergleichsweise sehr niedrigen Temperatur von 41°C wächst (Schnabel et al., 1982). Diese Diskrepanz zwischen Schmelzpunkt der Virus-DNA und der Wachstumstemperatur des Wirts deutet auf einem sehr effizienten Temperaturschutz der DNA hin. Die Aminosäuresequenz der Strukturproteine von TTV1 weist aber in dieser Hinsicht keine Besonderheiten auf, so daß der Schutz der DNA erst durch das Zusammenspiel aller am Virusaufbau beteiligten Komponenten zustande kommt. Daß dieses Zusammenwirken der Virusbausteine eine außerordentliche Temperaturstabilität bewirken kann, zeigt sich auch bei dem *Thermoproteus tenax* Virus TTV4, das selbst nach sechzigminütigem Autoklavieren bei 120°C noch infektiös bleibt (M. Rettenberger, pers. Mitteilung).

Aufbau des Genoms

Auf dem Genom von TTV1 wurden eine Anzahl offener Leserahmen entdeckt. Auffallend dabei ist die ungleichmäßige Verteilung der offenen Leserahmen auf dem Genom. Während am Ende der Sequenz, im ClaI-Fragment Nr. 4 nur zwei offene Leserahmen vorkommen, sind es im ClaI-Fragment Nr. 1 auf einer gleich langen DNA-Sequenz 15 offene Leserahmen. Da die ClaI-Fragmente Nr. 4 und Nr. 6 in die Variantenbildung involviert sind, läßt sich wahrscheinlich damit auch die geringe Anzahl der offenen Leserahmen in diesen beiden Fragmenten erklären. Der konservierte Teil des Genoms enthält 84% aller offenen Leserahmen, davon konnten drei als Gene für die Strukturproteine TP1, TP2 und TP3 indentifi-

ziert werden. Diese drei Gene sind, soweit sich dies ohne Transkriptionsdaten sagen läßt, nicht in gemeinsamen Transkriptionseinheiten verbunden, da ihr Abstand voneinander relativ groß ist und zwischen ihnen noch weitere offene Leserahmen liegen, deren Funktion allerdings unbekannt ist. Polycistronische Transkriptionseinheiten sind eigentlich ein typisches Merkmal von Eubakterien, kommen aber auch bei Archaeobakterien vor. Eine enge Bindung proteinkodierender Gene wurde in methanogenen Archaeobakterien gefunden (Konheiser *et al.*, 1984; Hamilton und Reeve, 1985; Reeve *et al.*, 1986), obwohl ein gemeinsames Transkript nur in einem Fall nachgewiesen werden konnte (Konheiser *et al.*, 1984). In letzter Zeit wurde auch im virusähnlichen Partikel SSV1 eine Operonstruktur von Proteingenen und damit ein gemeinsames Transkript entdeckt (Reiter *et al.*, 1987), so daß man auch in TTV1 gemeinsame Transkripte von vornherein nicht ausschließen kann.

Ribosomenbindungsstellen und Promoterstrukturen

Die Konservierung der 3'-Enden der 16S rRNA zwischen Eubakterien und Archaeobakterien (Steitz, 1978) und die erfolgreiche Translation von Genen aus methanogenen Archaeobakterien in Eubakterien (Wood *et al.*, 1983; Bollschweiler und Klein, 1982) weist darauf hin, daß die Initiation der Translation in beiden Gruppen von Organismen durch den gleichen Mechanismus kontrolliert wird. Die Signalstruktur für diesen Mechanismus, eine kurze, dem 3'-Ende der 16S rRNA komplementäre Sequenz auf der mRNA, ist normalerweise einige

Basen oberhalb des Initiationskodons lokalisiert (Shine und Dalgarno, 1974). Im Falle von TTV1 konnte weder vor den Genen für die Strukturproteine, noch vor anderen offenen Leserahmen eine zum 3'-Ende der 16S rRNA von *Thermoproteus tenax* (CCUCCACUA; Leinfelder et al., 1985) komplementäre Sequenz gefunden werden. Da bis jetzt keine Daten über Gene von Proteinen im Wirt *Thermoproteus tenax* vorliegen, kann zu diesem Zeitpunkt nicht gesagt werden, ob dieser Befund verallgemeinert werden kann. Es ist durchaus möglich, daß auf Grund der hohen Wachstumstemperatur und des niedrigen GC-Gehalts der TTV1-DNA die Initiation nach einem anderen Schema verfährt. Allerdings scheint das dann auf die *Thermoproteales* beschränkt zu sein, da in den weniger thermophilen *Sulfolobales*, der anderen Gruppe dieses Zweiges der Archaeobakterien, mögliche Ribosomenbindungsstellen oberhalb der Initiationskodons von Proteingenen gefunden wurden (Reiter et al., 1987) und auch in methanogenen Archaeobakterien solche Bindungsstellen nachgewiesen werden konnten (Bollschweiler et al., 1985; Cue et al., 1985; Hamilton und Reeve, 1985).

Ein Vergleich von stark konservierten Expressionssignalen in den vier Gattungen *Thermoproteus*, *Sulfolobus*, *Halo-bacterium*, und *Methanococcus* führte zu einer Konsensussequenz für Promotoren (Reiter, 1988). Ein Teil dieser Sequenz, $\text{TTTA}^{\text{T}}_{\text{A}}\text{--20-25b--}^{\text{T}}_{\text{A}}\text{TC}^{\text{A}}_{\text{G}}$ (Reiter, 1988), ähnelte dabei der eukaryotischen TATA-Box (Konsensus: TATAAA). Im Falle von *Thermoproteus tenax* wurden Konsensussequenzen für Promotoren vor stabilen RNA-Genen gefunden (Wich et al,

1987), dabei war die Sequenzkonservierung in der Umgebung des Transkriptionsstarts auf ein GC-Dinukleotid beschränkt, das zudem in einem relativ variablen Abstand zum Initiationspunkt auftrat. Leider existieren keine Daten über Proteingene in *Thermoproteus*, so daß kein Vergleich mit den "upstream" Regionen der Proteingene in TTV1 möglich ist. Da auch keine Transkriptionsstarts kartiert sind, beschränkt sich der Vergleich der möglichen Promotorstrukturen auf die Konsensussequenz. Ein Teil dieser Sequenz, die "box A", findet sich jeweils vor den Genen für die Proteine TP1, TP2 und TP3, fehlt aber, sowohl im Wildtyp als auch in der Variante VT3, vor dem Gen für TPX. Es wird sich aber erst durch Kartierung von Transkriptionsstarts herausstellen, ob diese Sequenzen in TTV1 überhaupt Promotorstrukturen sind. Wie bei den Ribosomenbindungsstellen besteht nun auch in den möglichen Promotorstrukturen keine große Ähnlichkeit mit SSV1, dem anderen Vertreter dieses Zweiges der Archaeobakterien, in dem in allen Fällen zwei konservierte Bereiche vor konstitutiven Transkripten gefunden wurden (Reiter, 1988).

Die genomischen Varianten

Änderungen im Restriktionsmuster der TTV1-DNA führten zum Nachweis von bisher fünf Virusvarianten, die sich durch strukturelle Veränderungen des Genoms voneinander unterschieden. Eine ähnlich hohe Variabilität wurde auch im Phagen Φ H von *Halobacterium halobium* nachgewiesen (Schnabel et al., 1982). Dort ist die Variabilität des Genoms aber

nicht nur auf den Phagen beschränkt, sondern auch bestimmte phänotypische Merkmale von *Halobacterium* selbst unterliegen einer hohen Mutationsrate. Solche phänotypischen Veränderungen konnten dagegen in *Thermoproteus* bisher nicht gefunden werden.

Die Bedeutung dieser häufigen strukturellen Veränderungen im Genom ist noch unklar. Da die natürliche Evolution durch Selektion in der Regel zum Verschwinden von Eigenschaften führt, die keinen Vorteil haben, ist die Möglichkeit, daß diese Veränderungen für Virus und Wirt ohne Bedeutung sind, nahezu ausgeschlossen. Diese häufigen Strukturänderungen könnten einerseits Ausdruck eines besonders aktiven Evolutionssystems sein oder aber sie werden gezielt zur Regulation der Genexpression eingesetzt. Zellen benutzen häufig Rekombinationsereignisse, um zwischen zwei alternativen DNA-Anordnungen zu wählen, die es erlauben, zwei verschiedene Proteine oder einen Satz von Proteinen zu exprimieren. Solche "Switch-Systeme" sind inzwischen in eubakteriellen und eukaryontischen Systemen nachgewiesen worden. Erstaunlicherweise reguliert dieser Mechanismus oft die Produktion von Proteinen, die an der Außenseite der Organismen erscheinen. Beispiele dafür sind das Schwanzprotein des Phagen Mu, das durch das invertierbare *gin* Segment reguliert wird (Kamp et al., 1979) und das Flagellenantigen von *Salmonella typhimurium* (Zieg et al., 1978). In einer sehr komplexen Weise wird auch die Konstruktion von Genen für Antikörper (Davis et al., 1980) und die Variation in den Oberflächenproteinen von

Trypanosomen (Borst und Cross, 1982) durch DNA-Rearrangements erreicht. Vielleicht wird die Rekombination deswegen in der Kontrolle der Oberflächenproteine benutzt, weil die Umschaltung absolut ist und sie dem Organismus erlaubt, ausschließlich ein Oberflächenantigen herzustellen. Dadurch vermeidet er jegliche Reaktion gegen Antikörper, die von dem infizierten Wirt gegen andere Antigene gerichtet sind. Welche dieser Möglichkeiten nun für TTV1 zutreffen, wird erst eine genauere Analyse und ein Vergleich aller Eigenschaften der Phagenvarianten zeigen.

Ein Modell der Variantenbildung

Botstein *et al.* entwickelten anhand der temperenten Phagen der *Enterobacteriaceae* ein Modell, nachdem das Produkt der Evolution nicht ein gegebener Phage ist, sondern austauschbare genetische Elemente, die eine ganze Phagenfamilie bilden können (Botstein, 1980; Campbell und Botstein, 1983). Diese genetischen Elemente, die Module, können dabei für einzelne Gene oder ganze Gensätze kodieren und haben dabei jeweils eine spezielle biologische Funktion wie z.B. Immunität, Insertion, Rekombination, Replikation, Lyse oder Struktur. Der Austausch solcher Module durch andere Elemente mit gleicher biologischer Funktion geschieht durch Rekombination unter den Mitgliedern der Phagenfamilie. Diese Rekombination ist ein "crossing-over" homologer Regionen, die zwischen nicht-homologen DNA-Regionen verteilt sind und als eine Art "Steckkontakt" zwischen den Modulen wirken. Dabei ist gleichgültig, ob die Homologie in

funktionellen Bereichen liegt oder nicht, es genügt, daß homologe DNA-Bereiche zwischen den auszutauschenden Modulpaaren existieren. Bei neuer Anordnung der Module kann dabei durchaus eine verbesserte biologische Funktion resultieren. Die bis jetzt im Labor beobachteten Modulaustausche haben eine vergleichbare oder sogar höhere Frequenz als die normale Mutationsrate, d.h. sie wirken schneller und sind flexibler als die Evolution durch lineare Vererbung (Botstein, 1980).

Dieses Modell läßt sich in gewisser Weise auch auf die Varianten von TTV1 anwenden. Die 8 bp Konsensussequenz ist der homologe Bereich, an dem das "crossing-over" stattfindet und der die eigentlichen Module, die in die Variantenbildung verwickelten DNA-Sequenzen, dadurch auf verschiedenste Weise anordnen kann. Im Gegensatz zum eigentlichen Modulmodell kodieren diese kurzen Sequenzen aber nicht für einzelne Gene oder gar ganze Gensätze, sondern werden in einen bereits bestehenden Leserahmen eingefügt und verändern diesen entsprechend ihrer Anordnung. Deshalb ist die Anwendung des Modulmodells für die Variantenbildung zwar ein interessanter Aspekt, aber aus den oben genannten Gründen wahrscheinlich nicht der gegebene Mechanismus.

Eine weitere Art, genomische Varianten zu bilden, wäre durch "site-specific recombination".

Die Rekombination zwischen zwei Stellen auf einem einzelnen DNA-Molekül hat zwei mögliche Konsequenzen, die davon abhängen, wie die interagierenden Stellen auf der DNA

orientiert sind. Die Rekombination entfernt entweder das dazwischen liegende Segment, oder sie invertiert es, wenn die Rekombinationsstellen entgegengesetzt orientiert sind. Wenn man davon ausgeht, daß die in die Variantenbildung verwickelte Konsensussequenz ACXCCTAC die Erkennungsstelle für die Rekombination ist, ist die Invertierung eines Segments nicht möglich, da alle diese Sequenzen in der gleichen Orientierung vorliegen. Auch das Restriktionsmuster der Varianten läßt keinen Schluß auf eine Inversion zu. Was die Deletierbarkeit von Segmenten durch Rekombination betrifft, läßt sich damit die Bildung der Varianten VT1, VT2 und VT5 erklären, aber nicht die der Varianten VT3 und VT4. Das bedeutet, daß der Mechanismus der "site specific recombination" nicht ganz ausgeschlossen werden kann, er aber nicht die Bildung aller Varianten erklärt.

Transposition von DNA-Sequenzen durch eine mögliche Transposase wäre dagegen ein Mechanismus, der fähig ist, die Bildung aller Varianten zu bewerkstelligen.

Transposons haben allgemein die Eigenschaft, Gene in neue Abschnitte auf der DNA zu bewegen, die mit den alten nicht in Verbindung stehen. Zwei Typen von Transposons konnten in Bakterien entdeckt werden, komplexe DNA-Abschnitte und relativ einfache, die auch Insertionselemente genannt werden. In *Halobakterien* konnten sowohl im Genom, als auch in den Plasmiden und im Phagen ϕ H IS-Elemente nachgewiesen werden (Betlach et al., 1983; Pfeiffer et al., 1981; Schnabel et al., 1982).

Für Transposons sind aber die DNA-Abschnitte, die die Variabilität in TTV1 bewirken, viel zu kurz. Sie enthalten außerdem keine "inverted repeats" an ihren Flanken und kodieren auch nicht für Proteine, Eigenschaften, die vielen Transposons gemein sind. Auch die Möglichkeit, daß die besagten DNA-Sequenzen in TTV1 Insertionselemente sind, ist unwahrscheinlich. Sie besitzen zwar mit der 8 bp Konsensussequenz eine ähnlich kurze Erkennungssequenz auf der DNA wie die Insertionselemente, aber es fehlt ihnen die Eigenschaft der IS-Elemente, diese Sequenz beim Inserieren in die neue Stelle auf der DNA zu verdoppeln. Außerdem können diese kleinen Elemente nicht die Information für ihre Transposition enthalten. Die Bildung der Varianten bei TTV1 scheint dagegen eher ein einfacher "cut and paste" Mechanismus zu sein. Die 8 bp Konsensussequenz fungiert dabei als Erkennungssequenz eines Enzyms, das wie ein Restriktionsenzym die DNA an dieser Stelle aufschneidet. Vielleicht übernimmt dieses hypothetische Enzym dann auch gleich den Transport des herausgeschnittenen DNA-Stücks und inseriert es nach dem gleichen Mechanismus entweder selbst oder mit Hilfe einer Ligase in der neuen Stelle auf der DNA. Dieses mögliche Enzym hätte dann die gleiche Funktion und den gleichen Mechanismus wie eine Transposase, mit der Ausnahme, daß es nicht auf den dem Transposon entsprechenden DNA-Sequenzen kodiert ist, sondern an einer anderen Stelle auf dem TTV1 Genom. Ein weiterer Punkt, der dieses Modell favorisiert, ist, daß sich, auf Grund der Erkennungssequenzen und des möglichen Restriktionsmechanismus, die

Elemente präzise in die neue Stelle auf der DNA einfügen, so daß das Leseraster erhalten bleibt. Dies ist dann besonders wichtig, wenn durch die Insertion der kurzen DNA-Sequenzen ein bereits bestehender Leserahmen so verändert wird, daß er ein ähnliches oder völlig neues Protein kodiert. Dazu ist allerdings auch eine korrekte Auswahl der Schnittstellen nötig und eine Unterscheidung der "bona fide" Konsensussequenzen von den "Pseudo-Konsensussequenzen". Ein ähnlicher Mechanismus der korrekten Auswahl von Schnittstellen findet sich auch beim spleißen von mRNA-Vorläufern (Maniatis und Reed, 1987).

Das Protein TPX

Das Genom eines Virus kodiert hauptsächlich für seine Hüllproteine. Zusätzlich werden meist ein oder mehrere neue Enzyme oder Hilfsproteine synthetisiert, um eine erfolgreiche Virusvermehrung zu gewährleisten oder aber auch, um die eigene Vermehrung zu unterdrücken. Zu dieser Sorte von Proteinen scheint nun das TTV1-Protein TPX zu gehören. Es ist kein integrales Virusprotein, wird aber in großer Menge von virus-tragenden Wirtszellen exprimiert.

Für die Funktion des Proteins TPX lassen sich mehrere Möglichkeiten diskutieren. Neue virusspezifische Proteine werden z.B. benötigt, um Vorläufermoleküle zu bilden. Im Phagen T₄ wird mit Hilfe eines phagenkodierten Enzyms die spezifische DNA-Base 5-OH-Methylcytosin gebildet (Wood und Edgar, 1973). Auch zur Viruszusammensetzung werden virus-

kodierte Proteine eingesetzt. Da dabei sehr viele Virusbausteine zusammengefügt werden und dementsprechend viele Reifungsproteine nötig sind, könnte eine solche Funktion auch die Menge des in den Zellen nachgewiesenen Proteins TPX erklären. Ein Beispiel für ein solches Reifungsprotein ist das Gen-D Produkt des Phagen ϕ X174. Es wird in infizierten Zellen von allen phagenkodierten Proteinen am häufigsten produziert und dient als Gerüstprotein beim Zusammenbau von Kapsomeruntereinheiten (Burgess und Denhardt, 1969; Fujisawa und Hayashi, 1977). Ein weiteres Argument für ein mögliches Gerüstprotein TPX sind die langen Pro-Thr Sequenzen und die daraus resultierende Struktur. Eine Änderung dieser Sequenzen verbessert dabei vielleicht die Funktion des Proteins und beschleunigt damit die Reifung des Virus. Das mag auch als Erklärung für die DNA-Umlagerungen in den Varianten dienen.

Eine weitere Möglichkeit für die Funktion von TPX ist die eines Repressors. TPX würde dabei an spezifische Stellen auf der Virus-DNA binden, die Initiation der korrespondierenden mRNA Moleküle blockieren und die Zellen im "carrier state" halten. Die Variation im kodierenden Gen könnte in diesem Fall eine Verbesserung der Repressorwirkung bedeuten. Repressormoleküle kommen in der Wirtszelle normalerweise nur in geringer Anzahl von maximal 200 Kopien (λ -Repressor; Ptashne, 1967). TPX muß aber auf Grund der starken Immunreaktion in wesentlich größerer Menge in der Zelle vorliegen, als für ein Repressorfunktion notwendig wäre.

Eine endgültige Erklärung für die Funktion des Proteins TPX und damit vielleicht auch für die Variantenbildung wird aber erst durch genetische Studien mit Hilfe eines noch zu entwickelnden Plaquetests möglich sein.

VI. LITERATUR

Bertani, G. & Baresi, L. (1986) Looking for gene transfer mechanism in methanogenic bacteria. In: "Archaeobacteria '85" (O. Kandler und W. Zillig, Hrsg.), Gustav Fischer Verlag, Stuttgart-New York.

Betlach, M., Pfeiffer, F., Friedman, J. & Boyer, H.W. (1983) Bacterio-opsin mutants of *Halobacterium halobium*. Proc. Natl. Acad. Sci. USA 80, 1416-1420.

Bolivar, F. & Backman, K. (1979) Plasmids of *Escherichia coli* as cloning vectors. Meth. Enzymol. 68, 245-267.

Bollschweiler, C. & Klein, A. (1982) Polypeptide synthesis in *Escherichia coli* directed by cloned *Methanobrevibacter arboriphilus* DNA. Zbl. Bakt. Hyg., I. Abt. Orig. C3, 101-109.

Bollschweiler, C., Kühn, R. & Klein, A. (1985) Non-repetitive AT-rich sequences are found in intergenic regions of *Methanococcus voltae*. EMBO J. 4, 805-809.

Borst, P. & Cross, G. (1982) Molecular basis for trypanosome antigenic variation. Cell 29, 291-303.

Botstein, D. (1980) A theory of modular evolution for bacteriophages. Ann. New York Acad. Sci. 80, 484-491.

Boyer, H.W. & Roulland-Dussoix, D. (1969) A complementation analysis of the restriction and modification of DNA

in *Escherichia coli*. J. Mol. Biol. 41, 459-472.

Brock, T.D., Brock, K.M., Belley, R.T. & Weiss, R.L. (1972) *Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch. Microbiol. 84, 54-68.

Burgess, A.B. & Denhardt, D.T. (1969) Studies on ϕ X174 proteins. I. Phage specific proteins synthesized after infection of *E. coli*. J. Mol. Biol. 44, 377-386.

Burnette, W.N. (1980) "Western Blotting": Electrophoretic transfer of proteins from SDS-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal. Biochem. 112, 195-203.

Campbell, A. & Botstein, D. (1983) Evolution of the lambdoid phages. In: Lambda II (R.W. Hendrix, Hrsg.) S. 365-380, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Chou, P.Y. & Fasman, G.D. (1978) Prediction of the secondary structure of proteins from there amino acid sequence. Adv Enzymol. 47, 45-148.

Cue, D., Beckler, G.S., Reeve, J.N. & Konisky, J. (1985) Structure and sequence divergence of two archaeobacterial genes. Proc. Natl. Acad. Sci. USA 82, 4207-4211.

Daniels, L.L. & Wais, A.C. (1984) Restriction and modification of halophage S45 in *Halobacterium*. Curr.

Microbiol. 10, 133-136.

Daniels, C.J., Gupta, R. & Doolittle, W.F. (1985) Transcription and excision of a large intron in the tRNA^{Trp} gene of an archaebacterium, *Halobacterium volcanii*. J. Biol. Chem. 260, 3132-3134.

Davis, M.M., Kim, S.K. & Hood, L. (1980) Immunoglobulin class switching: Developmentally regulated DNA rearrangements during differentiation. Cell 22, 1-2.

Dennis, P.P. (1986) Molecular biology of archaebacteria. J. Bact. 168, 471-478.

Devereux, J., Haeberli, P. & Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12, 387-395.

Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.E., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N., & Woese, C.R. (1980) The phylogeny of procaryotes. Science 209, 457-463.

Fujisawa, H. & Hayashi, M. (1977) Assembly of bacteriophage ϕ X174: Identification of virion capsid precursor and proposal of a model for the functions of bacteriophage gene products during morphogenesis. J. Virol. 24, 303-313.

Grant, W.N. & Williams, K.L. (1983) Monoclonal antibody

charakterisation of slime sheath: the extra-cellular matrix of *Dictyostelium discoideum*. EMBO J. 2, 935-940.

Hamilton, P.T. & Reeve, J.N. (1985) Structure of genes and an insertion element in the methane producing archaeobacterium *Methanobrevibacter smithii*. Mol. Gen. Genet. 200, 47-59.

Hilpert, R., Winter, J., Hammes, W. & Kandler, O. (1981) The sensitifity of archaeobacteria to antibiotics. Zbl. Bakt. Hyg., I. Abt. Orig. C2, 11-20.

Huet, J., Sentenac, A. & Fromageot, P. (1982) Spot-immunodetection of conserved determinants in eukaryotic RNA polymerases. J. Biol. Chem. 257, 2613-2618.

Huet, J., Schnabel, R., Sentenac, A. & Zillig, W. (1983) Archaeobacteria and eucaryotes possess DNA-dependent RNA polymerases of a common type. EMBO J. 2, 1291-1294.

Inouye, M. (1983) Experimental manipulation of gene expression. (Inouye, Hrsg.), Academic Press

Janekovic, D., Wunderl, S., Holz, I., Zillig, W., Gierl, A. & Neumann, H. (1983) TTV1, TTV2, and TTV3, a family of viruses of the extremely thermophilic, anaerobic, sulfur reducing archaeobacterium *Thermoproteus tenax*. Mol. Gen. Genet. 192, 39-45.

Kaine, B.P., Gupta, R. & Woese, C.R. (1983) Putative introns in tRNA genes of procaryotes. Proc. Natl. Acad. Sci. USA 80, 3309-3312.

Kaine, B.P. (1987) Intron-containing tRNA genes of *Sulfolobus solfataricus*. J. Mol. Evol. 25, 248-254.

Kamp, D., Chow, L.T., Broker, T.R., Kwoh, D., Zipser, D. & Kahman, R. (1979) Site-specific recombination in phage Mu. Cold Spring Harbor Symp. Quant. Biol. 43, 1159-1167.

Kandler, O. (1979) Zellwandstrukturen bei Methan-Bakterien. Zur Evolution der Prokaryonten. Naturwissenschaften 66, 95-105.

Kessel, M. & Klink, F. (1980) Archaeobacterial elongation factor is ADP-ribosylated by diphtera toxin. Nature 287, 250-251.

Kieny, M.P., Lathe, R. & Lecocq, J.P. (1983) New versatile cloning and sequencing vectors based on bacteriophage M13. Gene 26, 91-99.

Kjems, J. & Garrett, R.A. (1985) An intron in the 23S ribosomal RNA gene of the archaeobacterium *Desulfurococcus mobilis*. Nature 318, 675-677.

Konheiser, U., Pasti, G., Bollschweiler, C. & Klein, A. (1984) Physical mapping of genes coding for two subunits of methyl CoM reductase component C of *Methanococcus voltae*. Mol. Gen. Genet. 198, 146-152.

Kushner, D.J. (1985) The Halobacteriaceae. In: The Bacteria, Band VIII (C.R. Woese & R.S. Wolfe, Hrsg.) Academic Press, S. 171-214.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T₄. *Nature* 227, 680-685.

Langworthy, T.A., Smith, P.F. & Mayberry, W.R. (1972) Lipids of *Thermoplasma acidophilum*. *J. Bacteriol.* 112, 1193-1200.

Langworthy, T.A., Mayberry, W.R. & Smith, P.F. (1974) Long-chain glycerol diether and polyol dialkyl glycerol triether lipids of *Sulfolobus acidocaldarius*. *J. Bacteriol.* 119, 106-116.

Langworthy, T.A. (1977) Long-chain diglycerol tetraethers from *Thermoplasma acidophilum*. *Biochim. Biophys. Acta* 487, 37-50.

Lechner, K. & Böck, A. (1987) Cloning and nucleotide sequence of the gene for an archaeobacterial protein synthesis elongation factor Tu. *Mol. Gen. Genet.* 208, 523-528.

Leinfelder, W., Jarsch, M. & Böck, A. (1985) The phylogenetic position of the sulfur-dependent archaeobacterium *Thermoproteus tenax*: Sequence of the 16S rRNA gene. *System. Appl. Microbiol.* 6, 164-170.

Lottspeich, F. (1985) Microscale isocratic separation of phenylthio-hydantoin amino acid derivatives. *J. Chromatogr.* 326, 321-327.

Maniatis, T., Fritsch, E.F. & Sambrook, I. (1982)

Molecular cloning. A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York

Maniatis, T. & Reed, R. (1987) The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature* 325, 673-678.

Marmur, J. & Doty, P. (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.* 5, 109-118.

Martin, A., Yeats, S., Janekovic, D., Reiter, W.-D., Aicher, W. & Zillig, W. (1984) SAV1, a temperate u.v.-inducible virus-like particle from the archaeobacterium *Sulfolobus acidocaldarius* isolate B12. *EMBO J.* 3, 2165-2168.

Matheson, A.T. & Yaguchi, M. (1982) The evolution of the archaeobacterial ribosome. *Zbl. Bakt. Hyg., I. Abt. Orig.* C3, 192-199.

Messing, J. & Vieira, J. (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* 19, 269-276.

Messing, J. (1983) New M13 vectors for cloning. *Methods Enzymol.* 101, 20-78.

Mirault, M.-E., Scherrer, K. & Hansen, L. (1971) Isolation of preribosomes from HeLa cells and their characterisation by electrophoresis on uniform and exponential-gradient-polyacrylamide gels. *Eur. J. Biochem.* 23, 372-386.

Pauling, C. (1982) Bacteriophages of *Halobacterium halobium*: isolation from fermented fish sauce and primary characterisation. Can. J. Microbiol. 28, 916-921.

Parker, R.C. & Seed, B. (1980) Two-dimensional gel electrophoresis "Sea Plaque" agarose dimension. Meth. Enzymol. 65, 358-363.

Pecher, T. & Böck, A. (1981) In vivo susceptibility of halophilic and methanogenic organisms to protein synthesis inhibitors. FEMS Microbiol. 10, 295-297.

Pfeifer, F., Weidinger, G. & Goebel, W. (1981) Genetic variability in *Halobacterium halobium*. J. Bacteriol. 145, 375-381.

Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. & Surrey, S. (1982) "Nonrandom" DNA sequence analysis in bacteriophage M13 by dideoxychain-termination method. Proc. Natl. Acad. Sci. USA Vol. 79, 4298-4302.

Prangishvili, D. (1986) DNA-dependent DNA polymerases of the thermoacidophilic archaeobacterium *Sulfolobus acidocaldarius*. Molekularnaja Biologiya 20, 477-488.

Ptashne, M. (1967) Specific binding of the λ phage repressor to λ DNA. Nature 214, 232-234.

Pühler, G., Leffers, H., Gropp, F., Lottspeich, F., Palm, P., Garrett, R. & Zillig, W. (1988) Archaeobacterial RNA polymerases exhibit eukaryotic sequences and eubacterial gene organisation. Science, eingereicht zur Publikation.

Reeve, J.N., Hamilton, P.T., Beckler, G.S., Morris, C.J. & Clarke, C.H. (1986) Structure of methanogen genes. Syst. Appl. Microbiol. 7, 5-12.

Reiter, W.D., Palm, P., Henschen, A., Lottspeich, L., Zillig, W. & Grampp, B. (1987) Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. Mol. Gen. Genet. 206, 144-153.

Reiter, W.D. (1988) Genexpression im Archaeobakterium *Sulfolobus*: Untersuchungen anhand des virusartigen Partikels SSV1 und der ribosomalen RNA-Gene. Dissertation, München.

Rigby, P.W.J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. 113, 237-251.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. & Roe, B.A. (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143, 161-178.

Schnabel, H., Zillig, W., Pfäffle, M., Schnabel, R., Michel, H. & Delius, H. (1982a) *Halobacterium halobium* phage ϕ H. EMBO J. 1, 87-92.

Schnabel, H., Schramm, E., Schnabel, R. & Zillig, W.

(1982b) Structural variability in the genome of phage ϕ H of *Halobacterium halobium*. Mol. Gen. Genet. 188, 370-377.

Schnabel, R., Thomm, M., Gerardy-Schahn, R., Zillig, W., Stetter, K.O. & Huet, J. (1983) Structural homology between different archaebacterial DNA-dependent RNA polymerases analyzed by immunological comparison of their components. EMBO J. 2, 751-755.

Shine, J. & Dalgarno, L. (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsens triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA 71, 1342-1346.

Staden, R. (1980) A new computer method for the storage and manipulation of DNA gel reading data. Nucl. Acids Res. 8, 3673-3694.

Steitz, J.A. (1978) Methanogenic bacteria. Nature 273, 10

Stetter, K.O. & Zillig, W. (1985) *Thermoplasma* and the thermophilic sulfur-dependent archaebacteria. In: "The Bacteria" (I.C. Gunsalus, J.R. Sokatch, L.N. Ornston, C.R. Woese und R.S. Wolfe, Hrsg.), Academic Press, Band VIII, S. 85-170.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Thomas, P.S. (1983) Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods

Enzymol. 100, 255-266.

Tornabene, T.G. & Langworthy, T.A. (1979) Diphytanyl and dibiphytanyl glycerol ether lipids of methanogenic archaeobacteria. *Science* 203, 51-53.

Tornabene, T.G., Langworthy, T.A., Holzer, G. & Oro, J. (1979) Squalenes, phytanes and other isoprenoids as major neutral lipids of methanogenic and thermoacidophilic "archaeobacteria". *J. Mol. Evol.* 13, 73-83.

Torsvik, T. (1982) Characterisation of four bacteriophages for *Halobacterium*, with special emphasis on phage Hs1. In: "Archaeobacteria" (O. Kandler, Hrsg.), S. 351, Gustav Fischer Verlag, Stuttgart-New York.

Torsvik, T. & Dundas, I.D. (1974) Bacteriophage of *Halobacterium salinarium*. *Nature* 248: 680-681.

Torsvik, T. & Dundas, I.D. (1978) Halophilic phage specific for *Halobacterium salinarium* str.1. In: "Energetics and Structure of Halophilic Microorganisms" (S.R. Caplan & M. Ginzburg, Hrsg.), S. 609-614, Elsevier-North Holland Biomedical Press, Amsterdam-New York.

Torsvik, T. & Dundas, I.D. (1980) Persisting phage infection in *Halobacterium salinarium* str.1. *J. Gen. Virol.* 47, 29-36.

Towbin, H., Staehlin, T. & Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications.

Proc. Nat. Acad. Sci. USA 76, 4350-4354.

Vogelsang-Wenke, H. & Oesterhelt, D. (1986) Halophage Φ N. In: "Archaeobacteria '85" (O. Kandler und W. Zillig, Hrsg.), Gustav Fischer Verlag, Stuttgart-New York

Vogelstein, B. & Gillespie, D. (1979) Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA 76, 615-619.

Voller, A., Bidwell, D.E. & Bartlett, A. (1979) The enzyme linked immunosorbent assay; Dynatech.

Wais, A.C., Kon, M., MacDonald, R.E. & Stollar, B.D. (1975) Salt-dependent bacteriophage infecting *Halobacterium cutirubrum* and *H. halobium*. Nature 256, 314-315.

Whitman, W.B. (1985) Methanogenic bacteria. In: The Bacteria, Band VIII (C.R. Woese & R.S. Wolfe, Hrsg.), Academic Press, S. 3-84.

Wich, G., Leinfelder, W. & Böck, A. (1987) Genes for stable RNA in the extreme thermophile *Thermoproteus tenax*: introns and transcription signals. EMBO J. 6, 523-528.

Woese, C.R. & Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc. Natl. Acad. Sci. USA 74, 5088-5090.

Woese, C.R., Magrum, L.J. & Fox, G.E. (1978) Archaeobacteria. J. Mol. Evol. 11, 245-252.

Woese, C.R. & Olsen, G.J. (1986) Archaeobacterial

phylogeny: Perspectives on the eukaryotic kingdoms. System. Appl. Microbiol. 7, 161-177.

Wood, A.G., Redborg, A.H., Cue, D.R., Whitman, W.B. & Konisky, J. (1983) Complementation of *argG* and *hisA* mutations of *Escherichia coli* by DNA cloned from the archaeobacterium *Methanococcus voltae*. J. Bacteriol. 156, 19-29.

Wood, W.B. & Edgar, R.S. (1973) Building a bacterial virus. Scientific American. In The chemical basis of life (P.C. Hanawalt und R.H. Haynes, Hrsg.) San Francisco, Freeman.

Zabel, H.-P., Fischer, H., Holler, E. & Winter, J. (1985) *In vivo* and *in vitro* evidence for eucaryotic α -type DNA polymerases in methanogens. Purification of the DNA polymerase of *Methanococcus vannielii*. System. Appl. Microbiol. 6, 111-118.

Zieg, J., Silverman, M., Hilmen, M. & Simon, M.I. (1978) The mechanism of phase variation. in The Operon (J.H. Miller und W.S. Reznikoff, Hrsg.) pp. 411-423, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Zillig, W., Stetter, K.O., Schäfer, W., Janekovic, D., Wunderl, S., Holz, I. & Palm, p. (1981) *Thermoproteales*: A novel type of extremely thermoacidophilic anaerobic archaeobacteria isolated from Icelandic Solfataras. Zbl. Bakt. Hyg., I. Abt. Orig. C 2, 205-227.

Zillig, W., Stetter, K.O., Prangishvilli, D., Schäfer,

W., Wunderl, S., Janekovic, D., Holz, I. & Palm, P. (1982) *Desulfurococcaceae*, the second family of the extremely thermophilic, anaerobic, sulfur-respiring *Thermoproteales*. Zbl. Bakt. Hyg., I. Abt. Orig. C 3, 304-317.

Zillig, W., Schnabel, R. & Stetter, K.O. (1985a) Archaeobacteria and the origin of the eucaryotic cytoplasm. Curr. Top. Microbiol. Immunol. 114, 1-18.

Zillig, W., Stetter, K.O., Schnabel, R. & Thomm, M. (1985b) DNA-dependent RNA polymerases of the archaeobacteria. In: The Bacteria, (C.R. Woese & R.S. Wolfe, Hrsg.), Academic Press, Band VIII, S. 499-524.

Zillig, W., Yeats, S., Holz, I., Böck, A., Rettenberger, M., Gropp, F. & Simon, G. (1986a) *Desulfurolobus ambivalens*, gen. nov., sp. nov., an autotrophic archaeobacterium facultatively oxidizing or reducing sulfur. System. Appl. Microbiol. 8, 197-203.

Zillig, W., Gropp, F., Henschen, A., Neumann, H., Palm, P., Reiter, W.-D., Rettenberger, M., Schnabel, H. & Yeats, S. (1986b) Archaeobacterial virus host systems. System. Appl. Microbiol. 7, 58-66.

Zillig, W., Holz, I., Klenk, H.-P., Trent, J., Wunderl, S., Janekovic, D., Imself, E. & Haas, B. (1987) *Pyrococcus woesei* sp. nov., an ultra-thermophilic marine archaeobacterium, representing a novel order, *Thermococcales*. System. Appl. Microbiol. 9, 62-70.

VII. ABKÜRZUNGSVERZEICHNIS

BSA	Rinderserumalbumin
DNA	Desoxyribonukleinsäure
T ₄ -DNA	Desoxyribonukleinsäure des Bakteriophagen T ₄
T ₅ -DNA	Desoxyribonukleinsäure des Bakteriophagen T ₅
RNA	Ribonukleinsäure
ATP	Adenosin-5'-triphosphat
CTP	Cytidin-5'-triphosphat
GTP	Guanosin-5'-triphosphat
TTP	Thymin-5'-triphosphat
EDTA	Ethylendiamintetraessigsäure
SDS	Natriumdodecylsulfat
Tris	Trishydroxymethylaminomethan
bis	N,N'-Methylenbisacrylamid
Ci	Curie
m.o.i.	multiplicity of infection (Infektionsdosis)
rpm	Umdrehungen pro Minute
(k)bp	(Kilo)basenpaare
(k)D	(Kilo)dalton
U	Units
IPTG	Isopropyl- β -D-Thiogalactopyranosid
X-Gal	5-bromo-4-chloro-3-indolyl- β -galactosid
M	Molar
TEMED	N,N,N',N'-Tetramethylethylendiamin

LEBENS LAUF

Name:	Horst Neumann
Geburtsdatum:	20.10.1955
Geburtsort:	Kaufbeuren
Schulen:	
1962 - 1966	Grundschule in Kaufbeuren
1966 - 1975	Gymnasium in Kaufbeuren
1975	Abitur
1975 - 1977	Bundeswehr
akademische Ausbildung:	
1977 - 1983	Studium der Biologie an der Ludwig-Maximilians- Universität München
Februar 1982	Diplom-Hauptprüfung
März 1982 - April 1983	Diplomarbeit unter Anleitung von Prof. Dr. W. Zillig am Max-Planck- Institut für Biochemie, Martinsried
Mai 1983 - Dezember 1987	Doktorarbeit unter Anleitung von Prof. Dr. W. Zillig am Max-Planck- Institut für Biochemie, Martinsried

